

STIC-Biotech/ChemLib

49087

From: Mertz, Prema
Sent: Thursday, August 16, 2001 11:22 AM
To: STIC-Biotech/ChemLib
Subject: 09/521,195

Please search SEQ ID NO:1, 3, with protein databases.

Thanks
Prema Mertz, Ph.D.
Art Unit 1646
Crystal Mall 1, Room 10E-01
United States Patent & Trademark Office
(703) 308-4229

CRFE

*Edward Hart
Technical Info Specialist
STIC / Biotech
CM1 12C14 Tel: 305-9293*

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 16, 2001, 13:55:18 ; Search time 20.69 Seconds
(without alignments)
548.346 Million cell updates/sec

Title: US-09-521-195-1
Perfect score: 2845
Sequence: 1 MRDYDEVIAFLGNGPQRL.....KKTRDSMETEENPKVLITAF 551

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA: *
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	702	24.7	555	3	US-08-501-572-3
2	702	24.7	555	3	US-09-040-444-3
3	687.5	24.2	556	3	US-08-501-572-1
4	687.5	24.2	556	3	US-09-040-444-1
5	680	23.9	553	3	US-08-501-572-2
6	680	23.9	553	3	US-09-040-444-2
7	630	22.1	537	2	US-08-647-397-2
8	234.5	8.2	494	2	US-09-031-392-5
9	234.5	8.2	494	2	US-09-299-549-5
10	222	7.8	492	2	US-08-355-844-3
11	222	7.8	492	5	PCT-US95-16126-3
12	216	7.6	493	2	US-09-031-392-10
13	216	7.6	493	4	US-09-299-549-10
14	213.5	7.5	524	2	US-08-928-692-12
15	196	6.9	500	2	US-09-031-392-7
16	196	6.9	500	4	US-09-299-549-7
17	181	6.4	509	2	US-09-031-392-6
18	181	6.4	509	4	US-09-299-549-6
19	174.5	6.1	584	2	US-08-928-692-13
20	172	6.0	383	2	US-09-031-392-3
21	172	6.0	383	4	US-09-299-549-3
22	169.5	6.0	488	2	US-08-928-692-11
23	168.5	5.9	488	2	US-08-928-692-10
24	162	5.7	563	2	US-09-031-392-2
25	162	5.7	563	4	US-09-299-549-2
26	158.5	5.6	109	2	US-08-647-397-4
27	156.5	5.5	534	2	US-09-031-392-4

28	156.5	5.5	534	4	US-09-299-549-4	Sequence 4, Appli
29	110	3.9	473	1	US-08-597-236-13	Sequence 13, Appli
30	110	3.9	473	1	US-08-746-682A-13	Sequence 13, Appli
31	110	3.9	548	3	US-08-903-139B-9	Sequence 9, Appli
32	109.5	3.8	834	2	US-08-677-734A-9	Sequence 9, Appli
33	109.5	3.8	834	2	US-08-677-734A-10	Sequence 10, Appli
34	108.5	3.8	1299	4	US-08-460-900C-62	Sequence 62, Appli
35	107	3.8	1285	2	US-08-540-406-6	Sequence 6, Appli
36	107	3.8	1285	3	US-08-656-055-6	Sequence 6, Appli
37	107	3.8	1285	4	US-08-954-668-6	Sequence 6, Appli
38	107	3.8	1285	5	PCT-US95-13233-6	Sequence 6, Appli
39	103.5	3.6	1286	4	US-09-268-140-3	Sequence 3, Appli
40	103	3.6	548	3	US-08-903-139B-28	Sequence 28, Appli
41	101	3.6	470	2	US-08-724-394A-10	Sequence 10, Appli
42	99	3.5	455	1	US-08-035-928-2	Sequence 2, Appli
43	98	3.4	547	3	US-08-903-139B-7	Sequence 7, Appli
44	98	3.4	550	4	US-08-637-823B-30	Sequence 30, Appli
45	97	3.4	366	1	US-08-554-612C-50	Sequence 50, Appli

ALIGNMENTS

RESULT 1

US-08-501-572-3

; Sequence 3, Application US/08501572
; Patent No. 6063623
; GENERAL INFORMATION:
; APPLICANT: Koepsell, Hermann
; APPLICANT: Grundeman, Dirk
; APPLICANT: Gorboulev, Valentin
; TITLE OF INVENTION: Transport of protein Which Effects The
; TITLE OF INVENTION: Transport of Cationic Xenobiotics and/or Pharmaceuticals,
; TITLE OF INVENTION: DNA Sequences Encoding It And Their Use.
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan,Henderson, Farabow, Garrett & Dunner
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/501,572
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Toohy, Kimberlin M
; REGISTRATION NUMBER: 35,391
; REFERENCE/DOCKET NUMBER: 02481.1453-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)408-4000
; TELEFAX: (202)408-4400
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 555 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-501-572-3

SEQUENCE UNPARISON 'A'

Query Match 24.7%; Score 702; DB 3; Length 555;
Best Local Similarity 33.7%; Pred. No. 4.9e-65;
Matches 187; Conservative 93; Mismatches 229; Indels 46; Gaps 14;

QY 5 DEVIATLGEWGPQRLIFF---LLSASIIPNGFNGSVVFLAGTPBHRCPVDPANLS-- 59




```

; TELEPHONE: (202)408-4000
; TELEFAX: (202)408-4400
; INFORMATION FOR SEQ ID NO: 3
; SEQUENCE CHARACTERISTICS:
; LENGTH: 555 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-040-444-3

```

Query Match 24.7%; Score 702; DB 3; Length 555;
Best Local Similarity 33.7%; Pred. No. 4.9e-65;
Matches 187; Conservative 93; Mismatches 229; Indels

5	QY	5	DEVIAFLGFWGPGFORLFF---LLSASITPNFGNGMSVVFAGITPEHRCRVPDANLS--59
6	Db	6	DDVLEHGHEHFKQKOFFLALLATFAP---IYGVIVFLGTFDHCRCSPGVAEISLR62
60	QY	60	SAWR-----NNSVPLRLRDGREVPHCSRYRLA-TIANFSALGLEPGRDVLGQLEQESC11313
63	Db	63	CGWSPAELNVTVPGPCPAGEASPROCRRYEVDWNGSTFDCVDPLASLDTNRSRLPLGPC12222
114	QY	114	LDGWFEESODVTLSTVVTWENMLVCEDNWKVPLTTSIFFVGVLLGSFVSQLSDRGKKNVL17373
123	Db	123	RDGWY--ETPGSSIVTEFNLVCANGMMLDLQSSVYVNGFFIGSMSIGIYIADRFGRKLCL18000
174	QY	174	FATMAVQTGCFSLQITGISMEMTFLVIVGMGOISNYVVAFIIIGTEILGKSVRIIFSTL23333
181	Db	181	LTVLINAAAGVMAISPTYTWMLIFRLIOGLVSKAGMLIGYLITEFVGGRYR--RTV23737
234	QY	234	GVC--TTFVANGVMLPLFENAFIRDWRMLLALITVPGVLCVPLWTFIPSPRWLISQRRFR29111
238	Db	238	GIFYQVATVGLLVLAGVAYALPHWRWLOFTVALPHEFFLLYVWCIPESPRWLISQNKNA29777
292	QY	292	EAEIIITQAAKMNNTAVPAVI-----FDSVELNPLKQOKAFIDLFPRTNIAIWTIMS34555
298	Db	298	EAMRIIHKIAKNKSLPASLQRLLEEETGKKLNP-----SFDLVVRTPQIRKHTMIL35111
346	QY	346	LLIWLMLTSVGVFALSUDAPNLHGD-AYLNCFLSALLEIPAYITAMLLURTLPXYITAAV40000
352	Db	352	MYNWTSSVLYOGLIMHH-GLAGONHYLDFFYSALVEFFPAAFMILIIDRIGRRYPMAAS41000
405	QY	405	LPWGGVLLFTQLVPDVPDYFLISGLVLMRGFITSAFSMLYVFTAELYPTTLVRNNAVGVT46666
411	Db	411	NMVAGAACLASVFIIPGDLQWLKIIISCLGRMGITMAYEIVCLVNAELXYPTFLRNLGVHIC47000
465	QY	465	STASRVGSIAPYPFYILGAYNRM-----LPYTWGSLVILGIFTLFFPSPSIGMTLPET51888
471	Db	471	SMCDIGGLITTFPIYI-----RUTNIWLELPVFGVLGVAGGALLLPETKGAUPEI52999
519	QY	519	LEQMOKYKWFNRSGKK533
526	Db	526	IBEAEENMORPKNE540

```

RESULT      3
US-08-501-572-1
; Sequence 1, Application US/08501572
; Patent No. 6063623
; GENERAL INFORMATION:
; APPLICANT: Koepsell, Hermann
; APPLICANT: Grundeman, Dirk
; APPLICANT: Gorboulev, Valentin
; TITLE OF INVENTION: Transport protein which Effects The
; TITLE OF INVENTION: Transport of cationic xenobiotics and\or Pharmaceuticals,
; TITLE OF INVENTION: DNA Sequences Encoding It And Their Use.
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSES: Finnegan, Henderson, Farabow, Garrett & Dunner
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington

```

RESULT 2
US-09-040-444-3
; Sequence 3, Application US/09040444
; Patent No. 6063766
; GENERAL INFORMATION:
; APPLICANT: Koepsell, Hermann
; APPLICANT: Grundeman, Dirk
; APPLICANT: Gorboulev, Valentin
; TITLE OF INVENTION: Transport protein Which Effects The
; TITLE OF INVENTION: Transport Of Cationic Xenobiotics And Pharmaceuticals,
; TITLE OF INVENTION: DNA Sequences Encoding It And Their Use.
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnedan, Henderson, Farabow, Garrett & Dunner, L.L.P.
; STREET: 1300 I Street, N.W., Suite 700
; City: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/040,444
; FILING DATE: March 18, 1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: O'Connor, Steven P
; REGISTRATION NUMBER: 41,225
; REFERENCE/DOCKET NUMBER: 2481.1453-01
; COMMUNICATION INFORMATION:

1000

1000

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: * August 16, 2001, 13:55:18 ; Search time 20.69 Seconds
(without alignments)
548.346 Million cell updates/sec

Title: US-09-521-195-1
Perfect score: 2845
Sequence: 1 MRDYDEVIAFLGFWGPFQRL.....KKTRDSMETEENPKVLITAF 551

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents.AA.*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	702	24.7	555	3 US-08-501-572-3	Sequence 3, Appli
2	702	24.7	555	3 US-09-040-444-3	Sequence 3, Appli
3	687.5	24.2	556	3 US-08-501-572-1	Sequence 1, Appli
4	687.5	24.2	556	3 US-09-040-444-1	Sequence 1, Appli
5	680	23.9	553	3 US-08-501-572-2	Sequence 2, Appli
6	680	23.9	553	3 US-09-040-444-2	Sequence 2, Appli
7	630	22.1	537	2 US-08-647-397-2	Sequence 5, Appli
8	234.5	8.2	494	2 US-09-031-392-5	Sequence 5, Appli
9	234.5	8.2	494	2 US-09-299-549-5	Sequence 5, Appli
10	222	7.8	492	2 US-08-353-844-3	Sequence 3, Appli
11	222	7.8	492	5 PCT-US95-16126-3	Sequence 3, Appli
12	216	7.6	493	4 US-09-031-392-10	Sequence 10, Appl
13	216	7.6	493	4 US-09-299-549-10	Sequence 10, Appl
14	213.5	7.5	524	2 US-08-928-692-12	Sequence 12, Appl
15	196	6.9	500	2 US-09-031-392-7	Sequence 7, Appli
16	196	6.9	500	4 US-09-299-549-7	Sequence 7, Appli
17	181	6.4	509	2 US-09-031-392-6	Sequence 6, Appli
18	181	6.4	509	4 US-09-299-549-6	Sequence 6, Appli
19	174.5	6.1	584	2 US-08-928-692-13	Sequence 13, Appl
20	172	6.0	383	2 US-09-031-392-3	Sequence 3, Appli
21	172	6.0	383	4 US-09-299-549-3	Sequence 3, Appli
22	169.5	6.0	488	2 US-08-928-692-11	Sequence 11, Appl
23	168.5	5.9	488	2 US-08-928-692-10	Sequence 10, Appl
24	162	5.7	563	4 US-09-031-392-2	Sequence 2, Appli
25	162	5.7	563	4 US-09-299-549-2	Sequence 2, Appli
26	158.5	5.6	109	2 US-08-647-397-4	Sequence 4, Appli
27	156.5	5.5	534	2 US-09-031-392-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1
US-08-501-572-3
; Sequence 3, Application US/08501572
; Patent No. 6063623

; GENERAL INFORMATION:
; APPLICANT: Koepsell, Hermann
; APPLICANT: Grundeman, Dirk
; APPLICANT: Gorboulev, Valentin
; TITLE OF INVENTION: Transport protein Which Effects The
; TITLE OF INVENTION: Transport Of Cationic Xenobiotics and/or Pharmaceuticals,
; TITLE OF INVENTION: DNA Sequences Encoding It And Their Use.
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Carrett & Dunner
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/501,572
; FILING DATE:
; CLASSIFICATION: 424

; ATTORNEY/AGENT INFORMATION:
; NAME: Toohey, Kimberlin M
; REGISTRATION NUMBER: 35,391
; REFERENCE/DOCKET NUMBER: 02481.1453-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)408-4000
; TELEFAX: (202)408-4400
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 555 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-501-572-3

Query Match 24.7%; Score 702; DB 3; Length 555;
Best Local Similarity 33.7%; Pred. No. 4.9e-65;
Matches 187; Conservative 93; Mismatches 229; Indels 46; Gaps 14;

QY 5 DEVIAFLGFWGPFQRLFF---LLSASIIPIGFMGMSVFLAGTPEHRCRVPDANLUS-- 59

```

;
; TELEPHONE: (202)408-4000
; TELEFAX: (202)408-4400
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 555 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-09-040-444-3

Db 6 DDVLEHGGEFFQKQMFELLALSATAP---IYVGIVFLGFTDHRCSRPGVAELSLR 62
QY 60 SAWR-----NNSVPLRLRDGREVPHSCSYRLA-TIANFSALGLEPGRDVDLGQLEQESC 113
Db 63 CGWSPAEELNYTPGPGAGEASPROCRRYEVDWVNOSTFDCVDPDLASLDTNRSRLPLGFC 122
QY 114 LDGWEFSQDYLSTVVTENLVNVCEDNWKVPLTSTLFFVGVLLGSFVSQGLSDRFGRKNVL 173
Db 123 RDGWY--ETPGSSIVTEFNLCANSWMLDLFOSSVNVGFFIGSMISGYIADRFGRKLCL 180
QY 174 FATMAVQTGFSFIQFSISWEMFTVLFVIVMGQISNVVAVFILGTEILGKSVRIIFSTL 233
Db 181 LTTVLINAAAGVLMASPTYTWMFLIRLQGLVSKAGWLGIVLITFEVGGRYR---RTV 237
QY 234 GVC--TFPFAVGVMLLPLFAFYFIRDRMMLLALTVPGLVCLVPLWFWFIPESPRWLISORRFR 291
Db 238 GIFQYAVTVGLVLVAGVAYALPHRWLQFTVALPNFFELLYWCIPESPRWLISQNKNA 297
QY 292 EAEDIIQKAAMNNTAVPAVI-----FDSVEELNPLKQKAFILDLFRTNIAIMTMS 345
Db 298 EAMRIIKHIAKNGKSLPASLQRLLEBETGKKNP-----SFLDLVRTQIRKHTMIL 351
QY 346 LLLWMLTSVGFYFALSADPNLHGD-AYLNCFLSALIEIPAYITAWLLRLTPRYIAAV 404
Db 352 MYNFTSSVLYQGLIMHM-GLAGDNIYLDFFYSALVEPPAFMIIILIDRIGRRYPWAAS 410
QY 405 LFWGGGVLLFIQLVPDYVYFSLGVLGKGTSAFMSLVYFAELVPTLVNMAVGV 464
Db 411 NMVAGAACLASVFIIPGDLQWLKIIISCLGRMGITMAYEIVCLVNAELYPTFIRNLGVHIC 470
QY 465 STASRVGSIIPFYVYLGAYNRM-----LPYIVMGSLTVLIGTIFLFPESLGMTLPET 518
Db 471 SSMCDIGGIITPELVY-----RLTNWLELPLMVFGVLVAGGLALLPETKGPALPET 525
QY 519 LEQMOKVKWFRSGKK 533
Db 526 IEAEENMQRPKNKE 540

RESULT 2
US-09-040-444-3
; Sequence 3, Application US/0904044
; Patent No. 6063766
; GENERAL INFORMATION:
; APPLICANT: Koepsell, Hermann
; APPLICANT: Grundeman, Dirk
; APPLICANT: Gorboulev, Valentin
; TITLE OF INVENTION: Transport protein which effects the
; TITLE OF INVENTION: Transport of cationic xenobiotics and/or pharmaceuticals,
; TITLE OF INVENTION: DNA sequences encoding it and their use.
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner, L.L.P.
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/040,444
; FILING DATE: March 18, 1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: O'Connor, Steven P
; REGISTRATION NUMBER: 41,225
; REFERENCE/DOCKET NUMBER: 2481.1453-01
; TELECOMMUNICATION INFORMATION:

```

```

;
; TELEPHONE: (202)408-4000
; TELEFAX: (202)408-4400
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 555 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-09-040-444-3

Query Match 24.7%; Score 702; DB 3; Length 555;
Best Local Similarity 33.7%; Pred. No. 4.9e-65;
Matches 187; Conservative 93; Mismatches 229; Indels 46; Gaps 14;

QY 5 DEVIATLGEWGPQRLIFF---LLSASIPNGFNGSVVFLAGTPEHRCRVDAANLS-- 59
Db 6 DDVLEHGGEFFQKQMFELLALSATAP---IYVGIVFLGFTDHRCSRPGVAELSLR 62
QY 60 SAWR-----NNSVPLRLRDGREVPHSCSYRLA-TIANFSALGLEPGRDVDLGQLEQESC 113
Db 63 CGWSPAEELNYTPGPGAGEASPROCRRYEVDWVNOSTFDCVDPDLASLDTNRSRLPLGFC 122
QY 114 LDGWEFSQDYLSTVVTENLVNVCEDNWKVPLTSTLFFVGVLLGSFVSQGLSDRFGRKNVL 173
Db 123 RDGWY--ETPGSSIVTEFNLCANSWMLDLFOSSVNVGFFIGSMISGYIADRFGRKLCL 180
QY 174 FATMAVQTGFSFIQFSISWEMFTVLFVIVMGQISNVVAVFILGTEILGKSVRIIFSTL 233
Db 181 LTTVLINAAAGVLMASPTYTWMFLIRLQGLVSKAGWLGIVLITFEVGGRYR---RTV 237
QY 234 GVC--TFPFAVGVMLLPLFAFYFIRDRMMLLALTVPGLVCLVPLWFWFIPESPRWLISORRFR 291
Db 238 GIFQYAVTVGLVLVAGVAYALPHRWLQFTVALPNFFELLYWCIPESPRWLISQNKNA 297
QY 292 EAEDIIQKAAMNNTAVPAVI-----FDSVEELNPLKQKAFILDLFRTNIAIMTMS 345
Db 298 EAMRIIKHIAKNGKSLPASLQRLLEBETGKKNP-----SFLDLVRTQIRKHTMIL 351
QY 346 LLLWMLTSVGFYFALSADPNLHGD-AYLNCFLSALIEIPAYITAWLLRLTPRYIAAV 404
Db 352 MYNFTSSVLYQGLIMHM-GLAGDNIYLDFFYSALVEPPAFMIIILIDRIGRRYPWAAS 410
QY 405 LFWGGGVLLFIQLVPDYVYFSLGVLGKGTSAFMSLVYFAELVPTLVNMAVGV 464
Db 411 NMVAGAACLASVFIIPGDLQWLKIIISCLGRMGITMAYEIVCLVNAELYPTFIRNLGVHIC 470
QY 465 STASRVGSIIPFYVYLGAYNRM-----LPYIVMGSLTVLIGTIFLFPESLGMTLPET 518
Db 471 SSMCDIGGIITPELVY-----RLTNWLELPLMVFGVLVAGGLALLPETKGPALPET 525
QY 519 LEQMOKVKWFRSGKK 533
Db 526 IEAEENMQRPKNKE 540

RESULT 3
US-08-501-572-1
; Sequence 1, Application US/08501572
; Patent No. 6063623
; GENERAL INFORMATION:
; APPLICANT: Koepsell, Hermann
; APPLICANT: Grundeman, Dirk
; APPLICANT: Gorboulev, Valentin
; TITLE OF INVENTION: Transport protein which effects the
; TITLE OF INVENTION: Transport of cationic xenobiotics and/or pharmaceuticals,
; TITLE OF INVENTION: DNA sequences encoding it and their use.
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington

```

```

; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/501,572
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Toohey, Kimberlin M
; REGISTRATION NUMBER: 35,391
; REFERENCE/DOCKET NUMBER: 02481.1453-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)408-4000
; TELEFAX: (202)408-4400
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 556 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-501-572-1

```

```

Query Match      24.28; Score 687.5; DB 3; Length 556;
Best Local Similarity 33.28; Pred. No. 1.6e-63;
Matches 184; Conservative 88; Mismatches 230; Indels 53; Gaps 11;

QY 1 MRDYDEVIAFLGEMGPFORLIFFLL---SASIIPIGFMGMSVVFLAGTPHRCRVPDAA 57
Db 1 MPTVDVLEQGEFGWFOKQAFLLCLLSASLAP---IYGVIFLGTPTGHCYQNPQVAE 57

QY 58 LSS--AWR-----NNSVP-LRLRGREVPHSCSYRL-----ATIANFSA 95
Db 58 LSQRGWSQAEELNVTVPGLPDSDEASFLSQCMRYEVDWNGSTLDCVDP LSSLVANRSQL 117

QY 96 GLEPRGVDLGOLEQESCLDGEWFSQDYVLTSTVVTENLVCEDNKVKPLTSLFFVGVLL 155
Db 96 GLEPRGVDLGOLEQESCLDGEWFSQDYVLTSTVVTENLVCEDNKVKPLTSLFFVGVLL 155

QY 118 PLGP-----CEHGMY--DTPGSSIVTEFNLCVGDANKVDLQSCVNLGFFL 162
Db 118 PLGP-----CEHGMY--DTPGSSIVTEFNLCVGDANKVDLQSCVNLGFFL 162

QY 156 GSFVSQSLSDREGRKNVLFATMAVOTGFSFLQIFSIWEMFTVLFIIVGMQISNYVAF 215
Db 156 GSFVSQSLSDREGRKNVLFATMAVOTGFSFLQIFSIWEMFTVLFIIVGMQISNYVAF 215

QY 163 GSVLVGYIADRGKRLCLLVTLVTSVSGVLTAVAPDYSMLLFRLLQGMVSKGWSYSGY 222
Db 163 GSVLVGYIADRGKRLCLLVTLVTSVSGVLTAVAPDYSMLLFRLLQGMVSKGWSYSGY 222

QY 216 ILGTEILGKSVR---IIFSTLGVCTFFAVGYMLLPLFAFIRDRWMLLLALTPVGVLCV 271
Db 216 ILGTEILGKSVR---IIFSTLGVCTFFAVGYMLLPLFAFIRDRWMLLLALTPVGVLCV 271

QY 223 TLITEFVSGYRRTTAYLYQMA-----FTVGLVGLAGVAYAIPOWRMLQLAVSLPTFL 277
Db 223 TLITEFVSGYRRTTAYLYQMA-----FTVGLVGLAGVAYAIPOWRMLQLAVSLPTFL 277

QY 272 PLWVFIPESPRWLISORPREAEDIIOKAAKNNTTAVPAVIFDSVEELNPLKQOKAFILD 331
Db 272 PLWVFIPESPRWLISORPREAEDIIOKAAKNNTTAVPAVIFDSVEELNPLKQOKAFILD 331

QY 278 LYWFVFPESPRWLLSQKRTTRAVRINEQIAQKNKVPVADLKLMLCLEDEASEKSPSFAD 337
Db 278 LYWFVFPESPRWLLSQKRTTRAVRINEQIAQKNKVPVADLKLMLCLEDEASEKSPSFAD 337

```

```

RESULT 4
US-09-040-444-1
; Sequence 1, Application US/09040444
; Patent No. 6063766
; GENERAL INFORMATION:
; APPLICANT: Koepsell, Hermann
; APPLICANT: Grundeman, Dirk
; APPLICANT: Gorboulev, Valentin
; TITLE OF INVENTION: Transport of cationic xenobiotics and/or pharmaceuticals,
; TITLE OF INVENTION: Transport of cationic xenobiotics and/or pharmaceuticals,
; TITLE OF INVENTION: DNA Sequences Encoding It And Their Use.
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner, L.L.P.
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/040,444
; FILING DATE: March 18, 1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: O'Connor, Steven P
; REGISTRATION NUMBER: 41,225
; REFERENCE/DOCKET NUMBER: 2481.1453-01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)408-4000
; TELEFAX: (202)408-4400
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 556 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-09-040-444-1

```

```

Query Match      24.28; Score 687.5; DB 3; Length 556;
Best Local Similarity 33.28; Pred. No. 1.6e-63;
Matches 184; Conservative 88; Mismatches 230; Indels 53; Gaps 11;

QY 1 MRDYDEVIAFLGEMGPFORLIFFLL---SASIIPIGFMGMSVVFLAGTPHRCRVPDAA 57
Db 1 MPTVDVLEQGEFGWFOKQAFLLCLLSASLAP---IYGVIFLGTPTGHCYQNPQVAE 57

QY 58 LSS--AWR-----NNSVP-LRLRGREVPHSCSYRL-----ATIANFSA 95
Db 58 LSQRGWSQAEELNVTVPGLPDSDEASFLSQCMRYEVDWNGSTLDCVDP LSSLVANRSQL 117

QY 96 GLEPRGVDLGOLEQESCLDGEWFSQDYVLTSTVVTENLVCEDNKVKPLTSLFFVGVLL 155
Db 96 GLEPRGVDLGOLEQESCLDGEWFSQDYVLTSTVVTENLVCEDNKVKPLTSLFFVGVLL 155

QY 118 PLGP-----CEHGMY--DTPGSSIVTEFNLCVGDANKVDLQSCVNLGFFL 162
Db 118 PLGP-----CEHGMY--DTPGSSIVTEFNLCVGDANKVDLQSCVNLGFFL 162

QY 156 GSFVSQSLSDREGRKNVLFATMAVOTGFSFLQIFSIWEMFTVLFIIVGMQISNYVAF 215
Db 156 GSFVSQSLSDREGRKNVLFATMAVOTGFSFLQIFSIWEMFTVLFIIVGMQISNYVAF 215

QY 163 GSVLVGYIADRGKRLCLLVTLVTSVSGVLTAVAPDYSMLLFRLLQGMVSKGWSYSGY 222
Db 163 GSVLVGYIADRGKRLCLLVTLVTSVSGVLTAVAPDYSMLLFRLLQGMVSKGWSYSGY 222

QY 216 ILGTEILGKSVR---IIFSTLGVCTFFAVGYMLLPLFAFIRDRWMLLLALTPVGVLCV 271
Db 216 ILGTEILGKSVR---IIFSTLGVCTFFAVGYMLLPLFAFIRDRWMLLLALTPVGVLCV 271

QY 223 TLITEFVSGYRRTTAYLYQMA-----FTVGLVGLAGVAYAIPOWRMLQLAVSLPTFL 277
Db 223 TLITEFVSGYRRTTAYLYQMA-----FTVGLVGLAGVAYAIPOWRMLQLAVSLPTFL 277

QY 272 PLWVFIPESPRWLISORPREAEDIIOKAAKNNTTAVPAVIFDSVEELNPLKQOKAFILD 331
Db 272 PLWVFIPESPRWLISORPREAEDIIOKAAKNNTTAVPAVIFDSVEELNPLKQOKAFILD 331

QY 278 LYWFVFPESPRWLLSQKRTTRAVRINEQIAQKNKVPVADLKLMLCLEDEASEKSPSFAD 337
Db 278 LYWFVFPESPRWLLSQKRTTRAVRINEQIAQKNKVPVADLKLMLCLEDEASEKSPSFAD 337

```

QY 332 LFRTRNIAIMTMSLLMMLTSVGYFALSALDAPNLHGDVYLNCFSLALIEIPAYITAWLL 391
| | | | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | :
Db 338 LFRTPNLRKHVILMLWFSQVAVYQGLIMHVGATGANLYLDFYSSLYVEFPAPFIILVT 397
| | | | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | :
QY 392 LRTLPRLRVIIIAVFWGGVLLFTQIVPDVYFSLISGLVLMGKFGITSAFSLYVFTABE 451
| | | | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | :
Db 398 IDRIIGRIYPIAASNLVTGAACLLMIFIPHELHMLNVTACLGRMGATVLEWCLVNAEL 457
| | | | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | :
QY 452 YPTLVRNMAVGTSTASRVSIAPYEVY-LGAYNRMLPYIVMGSITVILGIFTLPFPES 510
| | | | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | :
Db 458 YPTFIRNLGMVMSALCDLGGIFTFPWFRLMEVWQALPILFSLGLTAGAMTLLLPET 517
| | | | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | :
QY 511 LGMLPETLEOMQKV 525
| | | | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | :
Db 518 KGVALPETIEEAHL 532
| | | | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | :

RESULT 5
US-08-501-572-2
; Sequence 2, Application US/08501572
; Patent No. 6063623
; GENERAL INFORMATION:
; APPLICANT: Koepsell, Hermann
; APPLICANT: Grundeman, Dirk
; APPLICANT: Gorboulev, Valentin
; TITLE OF INVENTION: Transport protein which effects the
; TITLE OF INVENTION: Transport of cationic xenobiotics and/or pharmaceuticals,
; TITLE OF INVENTION: DNA sequences encoding it and their use.
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: US/08/501,572
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Toohy, Kimberlin M
; REGISTRATION NUMBER: 35,391
; REFERENCE/DOCKET NUMBER: 02481.1453-00000
; TELEPHONE: (202)408-4000
; TELEFAX: (202)408-4400
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 553 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-501-572-2

Query Match 23.9%; Score 680; DB 3; Length 553;
Best Local Similarity 33.4%; Pred. No. 9.9e-63;
Matches 192; Conservative 85; Mismatches 228; Indels 70; Gaps 15;
QY 1 MRDYDEVIATLFGKPPQR---LIFFLLSASIIIPNGFNGMSVVFLAGTPEHRCRVDAAN 57
| | | | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | :
Db 1 MPTVDDILEQVSGSCWQKAFLLCLLSAFAF---ICVGIIVFLGTDPDHCQSPGVAE 57
| | | | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | :
QY 58 LSS-AWR-----NNSVPLRLRQREVPHSCSYR-----LATTA-NPSALG 96
| | | | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | :

Db 58 LSQRCGWSPAEELNYTPGLGPAGEAFLGQCRRYEVDWQNSALSQVDPLASLATNRSHLP 117
QY 97 LEPGRVDLDGLEQESCDGWESQDVYLSVTVTENWLNVEDNWKVPLTTSLLFFVGVLLG 156
| | | | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | :
Db 118 LGP-----CODGWY--DTPGSSIVTENLVNVCADSKLDDLQSCNLNAGFFFG 162
| | | | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | :
QY 157 SFVSGQLSDRFGRKNVLFATMAVOTGFSFIQISWEMPTVLFVIVGMQISNYVVAFI 216
| | | | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | :
Db 163 SLGVGYEADREFGRKLCLLGLTVLNAVSGVLMFSPNYMSMLLFRLLQGLVSKGNMAGYT 222
| | | | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | :
QY 217 LGTEILG----KSVRIIFSTLGVCTFAVGYMLLPFAFYFIRDRMLLLALTVPGLVCP 272
| | | | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | :
Db 223 LITEFVSGSRRTVAIMQMA----FTVGLVALTGLAYALPHWRMLQLAVSLPFLFL 277
| | | | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | :
QY 273 LWFPIESPRLWISQRRFREAEDIIKAAKMNNTAVPAVI-----FDSVEELNPLKQOK 326
| | | | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | :
Db 278 YYWCVPESPRLWSQKRNTAERKIMDHIAQKNGKLPADLKMLSLDEEDVTEKLS- 332
| | | | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | :
QY 327 AFILDLFRTRNIAIMTMSLLMMLTSVGYFALSALDAPNLHGDVYLNCFSLALIEIPAYI 386
| | | | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | :
Db 333 -SFADLFRTPRLKRTFILMLWFTDVLQGLILHMGATSGNLYLDFLYSALVEIPGAF 391
| | | | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | :
QY 387 TAWLLRLTPRRYIIAAVFWGGVLLFIOLVPDVFYFSLISGLVLMGKFGITSAFSLYV 446
| | | | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | :
Db 392 IALITIDRVGRIPYMAVSNLLAGAACLVIFISP-DLHNLNIIIMCVGRGITIAIQMICL 450
| | | | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | :
QY 447 FTABEYPTLVRNMAVGTSTASRVSIAPYEVY-LGAYNRMLPYIVMGSITVILGIFTL 505
| | | | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | :
Db 451 VNAELYPTFVRNLRVMVSCSLCDIGGIITPFIIVRLREVMQALPLILFAVLGLLAAGVTL 510
| | | | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | :
QY 506 FFPESLGMTLPETLEOMQKVWFSRSGKTKRDSMET 540
| | | | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | :
Db 511 LLPETKGALPETMKDAENL----GRKAKPKENT 540
| | | | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | :

RESULT 6
US-09-040-444-2
; Sequence 2, Application US/09040444
; Patent No. 6063766
; GENERAL INFORMATION:
; APPLICANT: Koepsell, Hermann
; APPLICANT: Grundeman, Dirk
; APPLICANT: Gorboulev, Valentin
; TITLE OF INVENTION: Transport protein which effects the
; TITLE OF INVENTION: Transport of cationic xenobiotics and/or pharmaceuticals,
; TITLE OF INVENTION: DNA sequences encoding it and their use.
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner, L.L.P.
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: March 18, 1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: O'Connor, Steven P
; REGISTRATION NUMBER: 41,225
; REFERENCE/DOCKET NUMBER: 2481.1453-01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)408-4000
; TELEFAX: (202)408-4400
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 553 amino acids

TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-040-444-2

Query Match 23.9%; Score 680; DB 3; Length 553;
Best Local Similarity 33.4%; Pred. No. 9.9e-63;
Matches 192; Conservative 85; Mismatches 228; Indels 70; Gaps 15;

QY 1 MRDVEIAFLGEGWPFQRLIFFLSASIIIPNGFMGMSVFLAGTPEHRCRVPDAA 57
Db 1 MPTVDDILEQVGSWFKQAFLILCLLSAFAF--ICVGIIVFLGFTPDHHCQSGVAE 57
QY 58 LSS--AMR-----NNSVPLRLRDGREVPHSCSRY-----LATIA-NFSALG 96
Db 58 LSQRCGSPAEELNYTPGLGPAGEAFLGQCRRYEVDMQNSALSCVDPLASLATNRSLP 117
QY 97 LEPRDVLGQLESCLDGWEFSQDYLSTVVTENNLYCEDNWKVPLTTSLEFFVGVLG 156
Db 118 LGP-----CQDGVVY--DTPGSSIVTEFNLCVADSKKLDLFOSCLNAGFFG 162
QY 157 SFVSQSLSDRFGKRNLFATMAVGTGFSFLQIFSIEMFTVLFVIVMGQISNYVVAFI 216
Db 163 SLGVGYFADREGRKCLLGTVLVNAVSVLMASFNTYMSMLLFRLLQGLVSKGNWAGYT 222
QY 217 LGTEILG----KSVRIIFSTLGVCTFFAVGYMLLPLFAYFIRDRMRLLLALTVPGLVCP 272
Db 223 LITEFVSGSRRTVAIMYQMA-----FTVGLVALTGLAYALPHWRWLQLAVSLPTFLFL 277
QY 273 LWWFIPESPRMLISQRRPREAEDIIQRAAKNNNTAVPAVI-----FDSVEELNPLKQOK 326
Db 278 YYWCVPESPRMLLSQKRNTAEIKIMDHIAQKNLPPADLKMLSLEEDVTEKLSF----- 332
QY 327 AFILDFRTRNIAITIMSLLLMWTSGVGFALSIDAPNLHGDAYLNCFLSALTEIPAYI 386
Db 333 -SFADLFTPLRKRTFLMLTFLWTFDSDVLYOGLIHMGATSGNLYLDFLYSALVEIPCAF 391
QY 387 TAWLLRLTPRYIIAAVLFWGGVLLFIQLVPVDDYVFLSLIGLVMKFGITSAFSMLYV 446
Db 392 IALITIDRVGRIPMAVSNLLAGAACLVIFISP-DLHWNLIIMCVGRMGITIAIQICL 450
QY 447 FTAELYPILVRNMAVGVSTASRVSIIATPFVY-LGAYNRMPLIYVNGSLTVLIGITL 505
Db 451 VNAELYPFVRNLRVMVSSLCDIGGIITPFRLEVVQALPLILFVGLLAAGVTL 510
QY 506 FPESIGLMTLPETLEOMOKVWFRSGKKTRDSMET 540
Db 511 LLPETGDALEPETMKDAENL-----GRKAKPKENT 540

RESULT 7
US-08-647-397-2
Sequence 2, Application US/08647397
Patent No. 5972702
GENERAL INFORMATION:
APPLICANT: Beier, David R.
APPLICANT: Brady, Kevin P.
TITLE OF INVENTION: OSTEOCLAST TRANSPORTER
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
STREET: 600 Atlantic Avenue
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/647,397
FILING DATE:
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Gates, Edward R.
REGISTRATION NUMBER: 31,616
REFERENCE/DOCKET NUMBER: B0801/7048
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-720-3500
TELEFAX: 617-720-2441
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 537 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-647-397-2

Query Match 22.1%; Score 630; DB 2; Length 537;
Best Local Similarity 30.0%; Pred. No. 1.7e-57;
Matches 165; Conservative 102; Mismatches 245; Indels 38; Gaps 12;

QY 4 YDEVIAFLGEGWPFQRLIFFLSASIIIPNGFMGMSVFLAGTPEHRCRVPDAA 63
Db 3 FSEILDRVSGMGPFQYLHVTLALPILGIANHNLQIFATTTPDHCRPPPNASL-EPW- 60
QY 64 NNSVPLRLRDGREVPHSCSRY---RLATANTALSALGLEPRDVLGQLESCLDGWEFS 120
Db 61 --VLPLG-PNGK--PEKCLRFVHLNPNASLPNDTOGATEP-----CLDGWIYN 102
QY 121 QDYLSTVVTENNLYCEDNWKVPLTTSLEFFVGVLGVSFGOLSDRFGKRNLFATMAVQ 180
Db 103 ST--RTVITENDVLGCGSNKLEMAQSVFMAGILVGGPVFGLSDRFGKRPILTWSYLL 160
QY 181 TGFSEFLQIFSIEMFTVLFVIVMGQISNYVVAFLITGLTEILGKSVRIIFS-TLGVCTFF 239
Db 161 AASGSSNAFSPSLTVYMIFFLGCSSISGLSLSLTIILNVWVPTSTRAISSTTIGYC--Y 218
QY 240 AVGYMLLPLFAYFIRDRMRLLLALTVPGLVCPVPLWVFIPESPWLISQRRPREAEDIIQK 299
Db 219 TIGQFILPLGLAYAVPQWRWLQLSVSAAFFIFLSLLSWVPESIRWLVLGKFSKALKTLQR 278
QY 300 AAKNN-----NTAPAVIFDSVEELNPLKQOKAFILDFRTRNIAITIMSLLLMWT 352
Db 279 VATFNGKKKEGKLTVEELAFNLOKDIITSK-VKYGSLDFRVSIILRRVTFCLSLAWFAT 337
QY 353 SVGYFALSIDAPNLHGDAYLNCFLSALTEIPAYITAWLLRLTPRYIIAAVLFWGGVLL 412
Db 338 GFAYISLAMGVVEFGVNIYILQIIFGGVDIPAKFITILSISYLGRRITQGFLLILAGVAI 397
QY 413 LFIQLVPVDDYVFLSLIGLVMKFGITSAFSMLYVFAELYPILVRNMAVGVSTASRVS 472
Db 398 LALIFVSSEMQLLRTALAVFGKCLSGFSCLFLYTSYLYPTVLRQTMGMSINWARGVS 457
QY 473 IAPYFVYLGAYNRMPLIYVNGSLTVLIGITLFFFPESLGMTLPETLEOMOKVWFRSGK 532
Db 458 MIAPLVKITGELQPFIPNVIFWTMLLGGSAAPFLETLNRPLEPIEDIIQ--DWIQTK 515
QY 533 KTRDSMETEE 542
Db 516 KTKQEPKAEK 525

RESULT 8
US-09-031-392-5
Sequence 5, Application US/09031392
Patent No. 5942398
GENERAL INFORMATION:
APPLICANT: Tartaglia, Louis A.
APPLICANT: Weng, Xun
TITLE OF INVENTION: NUCLEIC ACID MOLECULES

```

; TITLE OF INVENTION: ENCODING GLUTEX AND USES THEREOF
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/031,392
; FILING DATE: 26-FEB-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Melkielejohn, Ph.D., Anita L.
; REGISTRATION NUMBER: 35,283
; REFERENCE/DOCKET NUMBER: 07334/072001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 494 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-031-392-5

```

```

Query Match      8.2%; Score 234.5; DB 2; Length 494;
Best Local Similarity 25.2%; Pred No. 4.4e-16;
Matches 115; Conservative 76; Mismatches 156; Indels 109; Gaps 23;

```

```

QY 133 NLVCEDNWKVP-----LTTSLFFVGVLLGSFVSGQLSDRFGRKN--VLFATMAV 179
Db 43 NYTLERSETPPSVLLTSLWSLSVAIFSVMGIGSVGLFVNFRGRNSMLVNLAI 102

QY 180 QTG--FSELIQFISWEMFTVLFVIGM--GOISNVYVAFILGTILGKSVRIIFST--- 232
Db 103 AGCGLMGFCKI--AESVEMLIILGRLLIIGLFCGLCTGFVPMYI--GEISPTALRGAFGLNQ 159

QY 233 LGVCTFFAVGYMLLPFLFAYFI-----RDWRMLLIALTVP--PGVLCVPLWFWFIPESRWLIQ 287
Db 160 LGI-----VIGILVAQIFGLKVLGTEDLWPLLGLFTILPAILIQCAALPCPESPRFLIN 215

QY 288 RRPRE-----AEDIQKAAMNNTAVPAVIFDSVEELNPLKQOKAFILDLFRT 336
Db 216 RKEEKAKELIQLRWGTEDVAQDIQEMKD-----ESMRMSQEKQVTVLELFRAP 264

QY 337 N-----IAIMTMSLLWMLTSVGYFALSADPNLHGDALNCLFSLALIEPAYIT--- 387
Db 265 NYRQPIIISIMQLSQSLSGINAVFYSTGI-----FKDAGVQEPVYATIGA 311

QY 388 -----AWLLRLTPRRYIIAAVLFWGG-----GVLLFIQLVPVDYI-----FLSIG- 428
Db 312 GVVNTIFTVWSVFLVERAGRRTL--HLIGLGMAFCSILMTISLLKDNYSWMSFCIGA 369

QY 429 -LVMLGKFGITSAF--SMLYVFTAEPLTVLRNMAVGVSTAS-----RVGSIAPFYVLG 482
Db 370 ILVFAVFEIGGG--PIPMFIVIAELFGGPRPAAVAGCSNWTSNFLVLGLLFTSATYLG 428

QY 483 AYNRLPYIVMGSLTVLIGIFTLF--FPESLGMTLPE 517
Db 429 AY-----VFIVTVLIVFWVTFEFKVPETGRTEEE 460

```

```

RESULT
US-09-299-549-5

```

```

; Sequence 5, Application US/09299549
; Patent No. 6136547
; GENERAL INFORMATION:
; APPLICANT: Tartaglia, Louis A.
; APPLICANT: Weng, Xun
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
; TITLE OF INVENTION: GLUTEX AND USES THEREOF
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/299,549
; FILING DATE: 26-APR-1999
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/031,392
; FILING DATE: 26-FEB-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Melkielejohn, Ph.D., Anita L.
; REGISTRATION NUMBER: 35,283
; REFERENCE/DOCKET NUMBER: 07334/072002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 494 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-299-549-5

```

```

Query Match      8.2%; Score 234.5; DB 4; Length 494;
Best Local Similarity 25.2%; Pred No. 4.4e-16;
Matches 115; Conservative 76; Mismatches 156; Indels 109; Gaps 23;

```

```

QY 133 NLVCEDNWKVP-----LTTSLFFVGVLLGSFVSGQLSDRFGRKN--VLFATMAV 179
Db 43 NYTLERSETPPSVLLTSLWSLSVAIFSVMGIGSVGLFVNFRGRNSMLVNLAI 102

QY 180 QTG--FSELIQFISWEMFTVLFVIGM--GOISNVYVAFILGTILGKSVRIIFST--- 232
Db 103 AGCGLMGFCKI--AESVEMLIILGRLLIIGLFCGLCTGFVPMYI--GEISPTALRGAFGLNQ 159

QY 233 LGVCTFFAVGYMLLPFLFAYFI-----RDWRMLLIALTVP--PGVLCVPLWFWFIPESRWLIQ 287
Db 160 LGI-----VIGILVAQIFGLKVLGTEDLWPLLGLFTILPAILIQCAALPCPESPRFLIN 215

QY 288 RRPRE-----AEDIQKAAMNNTAVPAVIFDSVEELNPLKQOKAFILDLFRT 336
Db 216 RKEEKAKELIQLRWGTEDVAQDIQEMKD-----ESMRMSQEKQVTVLELFRAP 264

QY 337 N-----IAIMTMSLLWMLTSVGYFALSADPNLHGDALNCLFSLALIEPAYIT--- 387
Db 265 NYRQPIIISIMQLSQSLSGINAVFYSTGI-----FKDAGVQEPVYATIGA 311

QY 388 -----AWLLRLTPRRYIIAAVLFWGG-----GVLLFIQLVPVDYI-----FLSIG- 428
Db 312 GVVNTIFTVWSVFLVERAGRRTL--HLIGLGMAFCSILMTISLLKDNYSWMSFCIGA 369

QY 429 -LVMLGKFGITSAF--SMLYVFTAEPLTVLRNMAVGVSTAS-----RVGSIAPFYVLG 482

```



```

;
; ORIGINAL SOURCE:
; ORGANISM: Hum
; FEATURE:
; NAME/KEY: Pep
; LOCATION: 1..
; OTHER INFORMAT
; OTHER INFORMAT
PCT-US95-16126-3

```

Query Match	7.8%;	Score 222;	DB 5;	Length 492;
Best Local Similarity	24.0%;	Pred. NO. 9e-15;		
Matches 117; Conservative	74;	Mismatches 175;	Indels 122;	Gaps 21;

QY	119	FSQDYLSTVVTWENLVCDENKWKVPLTSLFFVGVLLGFSVQSOLSDRFRGR-----NV	172
Db	52	YGESILPTTLTTLWS-----LSVAIFSVMGIGFSVGLFVFNRFNRNSMLMML	101
QY	173	LFATMAVQTGSFSLQIFSIWEM-----FTVLFTVFMGQISNVVAFILGT	219
Db	102	LAFVASVLMGFSKL--GKSEFEMILGRFTIIGVYCGLTGTFVPMYGVSPAFRGALGT	158
QY	220	EILGKSVRIPTSLGVCTTFPAGVYMLLPFAYFI-----RD-WRMLLLATVPGVLCVPLW	274
Db	159	-----LHQLGI-----VVGILIAQVGLDSIMGNDKMLPLLLSIIFIPALLQCI	204
QY	275	WFIPESPRWLISORFREAEIDIQAKAKMNNTAVPAVIFDSVEELNPLAQO-----K	326
Db	205	PFCPESPRFLINNEENRAKSVLK-KLRGTA-----DVTHDLOEMKEESQMMREK	256
QY	327	AFILDLPRTN-----IATMTMSLLMLTSVGVFALSLDAPNLHGDAYLNCFLSALI	380
Db	257	VTILEFRSPAYROPILIAVVLQSQLSGINAVFYTSTI-----FEKAGV	303
QY	381	EIPAYIT-----AWLLRTLPRR--YIIAAVLFWGGGVLLFIQLVPVDYY--	423
Db	304	QQPVATIGSGIVNTAFTVVSLEFVVERAGRRTLHLGLAGMAQAILMTIALLEQLPW	363
QY	424	--FLSLGLVMGLKFGTISAFS-----MLYVFTAEYPTLVRNMNAVGVTSARVGS-IFA	475
Db	364	MSYLSIVAI---FGVAFGEVGPGRIPWFIVAELESQGPRAAIAVAGFSNWTSNFI	419
QY	476	PYFWYLGAYNRMLPYIVMGSLSLTVLIGIFTLF-FPESLGMNTLPETLEOMKVKWFRSG	534
Db	420	MCFQYVEQLCGPVYFIIFTVLLVLFPIRTYFKVPETKGRTFDEI-----ASG	473
QY	535	RDSMETEE	542
Db	474	OSDKTPPE	481

```

RESULT 12
US-09-031-392-10
; Sequence 10, Application US/09031392
; Patent No. 5942398
; GENERAL INFORMATION:
; APPLICANT: Tartaglia, Louis A.
; APPLICANT: Weng, Xun
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES
; TITLE OF INVENTION: ENCODING GLUTEX AND USE
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FASTSEQ for Windows Version 2.0

```

```

: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/031,392
: FILING DATE: 26-FEB-1998
: ATTORNEY/AGENT INFORMATION:
: NAME: Meiklejohn, Ph.D., Anita L.
: REGISTRATION NUMBER: 35,283
: REFERENCE/DOCKET NUMBER: 07334/072001
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 617/542-5070
: TELEFAX: 617/542-8906
: TELEX: 200134
: INFORMATION FOR SEQ ID NO: 10:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 493 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-09-031,392-10

```

Query Match 7.68; Score 216; DB 2; Length 493;

Best Local Similarity 24.2%; Pred. No. 3.8e-14;
Matches 111; Conservative 78; Mismatches 171; Indels 98; Gaps 19;

Qy	144	LTTSLLFPVGVLLGSFVSGQLSDRGRKNVLPA--TNVQTGFSP-LQIFSIWEMFTVLF	200
Db	68	LSVIFAVGGMISFLGXIGNRLGRKXAMLVNVLAIAGGLMLGAKXAXSEPMILGR	127
Qy	201	VIVGM-----GOISNVVAFILGT-EILGKSVRIIFS-TLGVCTFEAVGYML	245
Db	128	FLILGYCGLSSGVVPMYGEISPTALRGALGTNLQGLGVIGILLIAQVLGLDSL--LGNES	185
Qy	246	LPLPAYFIRDRMILLALTVPGVLCVPLMWFIPESPRLVI-----SQRREFRAE	294
Db	186	L-----WPLLGLGTVPALLQLLLPFCPESPRYLLINKNEEARAKKALQRLGTA	236
Qy	295	DIIOKAAMNNTAVPAVIFDSVEELNPLKQOKAIFLDLRTN-----IATMTNSLLL	348
Db	237	DVSGEVAEMD-----ESRMMXSEKXVSVELFNSRYRQPVIIAIVQLSQQL	285
Qy	349	WMLTSVGYFALSADPNLHGDYALNCFLSALIEIPAYIT-----AWLLRLT	395
Db	286	SGINAVEFYSTSI-----FEKAGVQPVVATIGAGVNTVTVVSVFVVERA	332
Qy	396	PRR--YIIAAVLFWGGVLLFIQLVPDY--YFLSLGLWMLKFGTSAFS-----MLXV	446
Db	333	GRTLHLGLLGGAGCAVLTIALALLDQVPMMSYYSIVAI--FGFVAFEPGPGIPWF	390
Qy	447	FTAEIYPTLVNNAVGVTSTASRVGS-ITAPYFVILGAYNRMPLPYMGLSLVLTIGIFT-	504
Db	391	IVALEFSQGRPRAIAVAGFSNWTSNFIVGLLEFQYIAELLGPVYFVFAVLLLPFIPTF	450
Qy	505	LFPFESIGMTLPETLEQMOKVWFRSGKKTROSMETEE	542
Db	451	LKYPETKGRFDFAAFAFRKXN--KXEOPEKSEIELE	486

```

RESULT 13
US-09-299-549-10
; Sequence 10, Application US/09299549
; Patent No. 6136547
; GENERAL INFORMATION:
; APPLICANT: Tartaglia, Louis A.
; APPLICANT: Weng, Xun
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
; TITLE OF INVENTION: GLUTEX AND USES THEREOF
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 16, 2001, 13:54:58 ; Search time 36.39 Seconds
(without alignments)
917.939 Million cell updates/sec

Title: US-09-521-195-1
Perfect score: 2845
Sequence: 1 MRDYDEVIAFLGEGWGFQRL.....KKTRDSMETENPKVLITAF 551

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues
Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_0601.*
1: /SID88/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SID88/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SID88/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SID88/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SID88/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SID88/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SID88/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SID88/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SID88/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SID88/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SID88/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SID88/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SID88/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SID88/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SID88/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SID88/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SID88/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SID88/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SID88/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SID88/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2845	100.0	551	AAV01649	A protein with cat
2	2470	86.8	553	AAV01851	A protein with cat
3	2470	86.8	553	AAE20579	Mouse OCTN1 amino
4	2218	78.0	557	AAV01650	A protein with cat
5	2218	78.0	557	AAV83929	Human carnitine tr
6	2164	76.1	557	AAV01652	A protein with cat
7	2164	76.1	557	AAV01652	Mouse OCTN2 amino
8	2164	76.1	557	AAV83930	Mouse carnitine tr
9	2034.5	71.5	564	AAE20578	Mouse OCTN3 protei
10	779.5	27.4	560	AAE43038	Human ORFX ORF2802
11	698	24.5	540	AAE49401	Murine organic ani

12	685.5	24.1	554	19	AAW64538	Human liver cell c
13	682	24.0	535	21	AAV51249	Rat liver anion tr
14	681.5	24.0	556	17	AAV77676	Rat OCT-1 protein.
15	653.5	23.0	548	21	AAE08823	A human organic an
16	651	22.9	542	21	AAV92902	Human cerebral org
17	650	22.8	551	21	AAE08824	A human organic an
18	637	22.4	536	21	AAV92903	Rat cerebral organ
19	636	22.4	561	18	AAW44196	Human organic anio
20	630	22.1	537	18	AAW44195	Mouse osteoclast t
21	624	21.9	607	21	AAE12131	Hydrophobic domain
22	606.5	21.3	545	22	AAE36553	Mouse organic anio
23	597.5	21.0	551	20	AAW88488	Rat organic anion
24	586	20.6	550	21	AAV44278	Human organic anio
25	583.5	20.5	563	20	AAW88489	Human organic anio
26	564	19.8	550	22	AAE69091	Human organic anio
27	417	14.7	480	21	AAE28044	Arabidopsis thalia
28	417	14.7	483	21	AAE28043	Arabidopsis thalia
29	412.5	14.5	521	21	AAE30885	Arabidopsis thalia
30	387.5	13.6	397	21	AAE30886	Arabidopsis thalia
31	382.5	13.4	515	21	AAE42522	Arabidopsis thalia
32	341	12.0	360	21	AAE30887	Arabidopsis thalia
33	325.5	11.4	378	21	AAE42523	Arabidopsis thalia
34	312.5	11.0	359	21	AAE42524	Arabidopsis thalia
35	305.5	10.7	439	22	AAE76766	Corynebacterium gl
36	303	10.7	297	21	AAE28045	Arabidopsis thalia
37	302.5	10.6	548	21	AAE44633	Human organic cati
38	277	9.7	339	21	AAE19406	Amino acid sequenc
39	272.5	9.6	231	17	AAE77677	Partial human OCT-
40	267	9.4	520	20	AAV06115	Human organic cati
41	267	9.4	520	21	AAV56881	Human saccharide-t
42	267	9.4	520	21	AAV53009	Human secreted pro
43	247	8.7	530	21	AAE29626	Cat flea HMT synap
44	243	8.5	399	22	AAE76717	Corynebacterium gl
45	243	8.5	492	12	AAE11360	Glucose Transporte

ALIGNMENTS

RESULT 1
AAV01649
ID AAY01649 standard; Protein; 551 AA.
XX
AC AAY01649;
XX
XX
DT 23-JUN-1999 (first entry)
XX
DE A protein with cation transporting activity.
XX
KW Organic cation transporter; OCT1; OCT2; drug development; fatty liver;
heart disease; cancer; anti-tumour drug; anticancer drug.
XX
OS Homo sapiens.
XX
PN WO913072-A1.
XX
PD 18-MAR-1999.
XX
PF 07-SEP-1998; 98WO-JP04009.
XX
PR 20-MAY-1998; 98JP-0156660.
PR 08-SEP-1997; 97JP-0260972.

(CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.

Nezu J, Oku A;

WPI: 1999-215062/18.

N-PSDB; AAX26879.

Genes homologous with organic cation transporters OCT1 and OCT2, useful in design of new drugs for treatment of diseases due to abnormality of the transporter functions

```

XX PS Claim 1: Page 41-45; 97pp; Japanese.
XX CC The present sequence represents a protein with cation transporting
XX CC activity. The genes are significantly homologous with organic cation
XX CC transporters OCT1 and OCT2. The genes may be used in drug development,
XX CC particularly in the treatment of diseases due to abnormality of the
XX CC organic cation transporter functions e.g. fatty liver, heart diseases
XX CC and cancers, by controlling such as by inhibition or activation.
XX CC Administration of anti-tumour and anticancer drugs in combination with
XX CC a transporter protein inhibiting agent allows the agents to penetrate
XX CC into the diseased cells to enhance the drug action.
XX SQ Sequence 551 AA;

Query Match 100.0%; Score 2845; DB 20; Length 551;
Best Local Similarity 100.0%; Pred. No. 1.2e-281;
Matches 551; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRDYDEVIAFLGEGWGPQRLIFFLLSASIIPIGNGFNGMSVVFVLAGTPEHRCRVPDAANLSS 60
DB 1 mrdydeviaflgwgpgfqrllifllsasiipngfngmsvfvlagtpehrcrvpdaanlss 60

QY 61 AWRNNSVPLRLRDGREVPHSCRYRLATIANFSALEPGSDRGVLDGQLEQESCLDGWEFS 120
DB 61 awrnnsvglrldgrevpshscryrlatiansalglepgrdvldgqlqescldgwefs 120

QY 121 ODVYLSVVTENLVCDNWKVPLTTSFFVGLGVSFVSGQLSDRGKKNVLFATMAVQ 180
DB 121 odvylstvtewnlvcdnwkvp ltsffvglgvsfvsgqlsdrfgrknvlfatmaq 180

QY 181 TGFSLQIFSIWEMFTLVFVIVGMQISNYVAVFVILGTEILGKSVRIIFSTLGVCTFFA 240
DB 181 tgfslqifswemftlvfivgmqisnyvavfvlgtelilgksvriifstlgtvctffa 240

QY 241 VGYMLPLFAYFIRDRWMLLALTVPGLCVPLWVFPESPRWLISQRRFRAEDIIQKA 300
DB 241 vgymlplfayfirdwrmlaltvpglcvplwvfpesprwlisqrrfraediika 300

QY 301 AKMNTAVPAVIFDSVEELNPLKQKAFILDLFRTNIAIMTMSLLIWLMLTSVGYFA 360
DB 301 akmntavpavifdsveelnplkqkafildlfrtniaimtmslllwmmltsvgyfals 360

QY 361 LDAPNLHGDAYLNCPLSALIEIPAYITAWLLRLPRYIIAANLVFWGGVLLFIQLVPV 420
DB 361 ldapnlhgdaylncplsalieipayitawllrlpryiaaavlfwggvllfiqlvpv 420

QY 421 DYVFLSIGVLMGLKFGITSAFSLMVLVFTAEYPTLVNMAVGVTSASRVGSIAPYFY 480
DB 421 dyvflsigvmlglkfgitsafslmvlvftaelyptlvnrmavgvtsasrvgsiapyfy 480

QY 481 LGAYNRLPYIVNGSLTVLIGITLFFPESLGMTLPETLEQMVKWPFPSCKKTRDSMET 540
DB 481 lgaynrmipyivngsltvligitlffpeslgmtlpetleqmkwkwpfscckktrdsmet 540

QY 541 EENPKVLITAF 551
DB 541 eenpkvlitaf 551

```

```

RESULT 2
AY01651
ID AAY01651 standard; Protein; 553 AA.
XX
XX AC AAY01651;
XX DT 23-JUN-1999 (first entry)
XX DE A protein with cation transporting activity.
XX KW Organic cation transporter; OCT1; OCT2; drug development; fatty liver;
KW heart disease; cancer; anti-tumour drug; anticancer drug.

```

```

XX OS Mus musculus.
XX PN WO9913072-A1.
XX PD 18-MAR-1999.
XX PF 07-SEP-1998; 98WO-JP04009.
XX PR 20-MAY-1998; 98JP-0156660.
XX PR 08-SEP-1997; 97JP-0260972.
XX PA (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
XX PI Nezu J, Oku A;
XX DR WPI; 1999-215062/18.
XX DR N-PSDB; AAX26898.
XX PT Genes homologous with organic cation transporters OCT1 and OCT2,
XX PT useful in design of new drugs for treatment of diseases due to
XX PT abnormality of the transporter functions
XX PS Claim 1; Page 63-67; 97pp; Japanese.
XX CC The present sequence represents a protein with cation transporting
XX CC activity. The genes are significantly homologous with organic cation
XX CC transporters OCT1 and OCT2. The genes may be used in drug development,
XX CC particularly in the treatment of diseases due to abnormality of the
XX CC organic cation transporter functions e.g. fatty liver, heart diseases
XX CC and cancers, by controlling such as by inhibition or activation.
XX CC Administration of anti-tumour and anticancer drugs in combination with
XX CC a transporter protein inhibiting agent allows the agents to penetrate
XX CC into the diseased cells to enhance the drug action.
XX SQ Sequence 553 AA;

```

```

Query Match 86.8%; Score 2470; DB 20; Length 553;
Best Local Similarity 84.6%; Pred. No. 2.3e-243;
Matches 468; Conservative 39; Mismatches 44; Indels 2; Gaps 1;

QY 1 MRDYDEVIAFLGEGWGPQRLIFFLLSASIIPIGNGFNGMSVVFVLAGTPEHRCRVPDAANLSS 60
DB 1 mrdydeviaflgwgpgfqrllifllsasiipngfngmsvfvlagtpehrcrvpdtvnllss 60

QY 61 AWRNNSVPLRLRDGREVPHSCRYRLATIANFSALEPGSDRGVLDGQLEQESCLDGWEFS 120
DB 61 awrnnsiplrldgrevpshscryrlatiansalglepgrdvldgqlqescldgweyd 120

QY 121 ODVYLSVVTENLVCDNWKVPLTTSFFVGLGVSFVSGQLSDRGKKNVLFATMAVQ 180
DB 121 odvylstvtewnlvcdnwkvp ltsffvglgvsfvsgqlsdrfgrknvlfatmaq 180

QY 181 TGFSLQIFSIWEMFTLVFVIVGMQISNYVAVFVILGTEILGKSVRIIFSTLGVCTFFA 240
DB 181 tgfslqifswemftlvfivgmqisnyvavfvlgtelilgksvriifstlgtvctffa 240

QY 241 VGYMLPLFAYFIRDRWMLLALTVPGLCVPLWVFPESPRWLISQRRFRAEDIIQKA 300
DB 241 vgymlplfayfirdwrmlaltvpglcvplwvfpesprwlisqrrfraediika 300

QY 301 AKMNTAVPAVIFDSVEELNPLKQKAFILDLFRTNIAIMTMSLLIWLMLTSVGYFA 358
DB 301 akmntavpavifdsveelnplkqkafildlfrtniaimtmslllwmmltsvgyfa 360

QY 359 LSDAPNLHGDAYLNCPLSALIEIPAYITAWLLRLPRYIIAANLVFWGGVLLFIQLV 418
DB 359 lsdapnlhgdaylncplsalieipayitawllrlpryiaaavlfwggvllfiqlv 420

QY 419 PDYVFLSIGVLMGLKFGITSAFSLMVLVFTAEYPTLVNMAVGVTSASRVGSIAPYFY 478
DB 421 pedyvflsigvmlglkfgitsafslmvlvftaelyptlvnrmavgvtsasrvgsiapyf 480

```

Qy 479 VYLGAYNRMPLPYIVMGSLTVLIGIFTLFFPESLGMPLPETLEOMQVKWFRSCKKTRDSM 538
Db 481 vylgaynrllpyilmslvtlglitlffesfgvtlpenleqmkgvrgcggkktvsv 540

Qy 539 ETEENPKVLITAF 551
Db 541 dreespkvlitaf 553

RESULT 3

AAB20579
ID AAB20579 standard; Protein; 553 AA.

XX AC

XX AC

XX DT 11-DEC-2000 (first entry)

XX DE Mouse OCTN1 amino acid sequence.

XX KW Mouse; transporter; OCTN1; OCTN2; OCTN3; organic cation transporter;
XX KW identification; regulator; carnitine transport.

XX OS Mus musculus.

XX PN WO200046368-A1.

XX PD 10-AUG-2000.

XX PF 04-FEB-2000; 2000WO-JP00619.

XX PR 05-FEB-1999; 99JP-0028406.

XX (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.

XX PA Nezu J, Ose A;

XX PI WPI; 2000-586982/55.

XX DR Organic cation transporter gene OCTN3 expressed in testis for
XX PT identification of regulators of carnitine transport for use as drugs

XX PS Example 2; Fig 2; 58pp; Japanese.

XX CC The present invention describes a mouse organic cation transporter
XX CC protein (OCTN3). Also described are: (1) a method for screening
XX CC compounds for their ability to regulate the transport of an organic
XX CC cation into the cell, by generating a cell expressing OCTN3 at the
XX CC cell membrane, contacting with the compound and organic cation, and
XX CC observing the degree of transport of the organic cation; and (2) a
XX CC method for screening compounds for their ability to be transported into
XX CC the cell by OCTN3, by generating a cell expressing OCTN3 at the cell
XX CC membrane, contacting with the compound and observing the degree of
XX CC transport of the compound. OCTN3 can be used for the identification of
XX CC regulators of the transport of organic cations (especially carnitine)
XX CC into cells by OCTN3, for use as drugs. The present sequence represents
XX CC the mouse OCTN1 amino acid sequence, which is used in an example from
XX CC the present invention.

XX SQ Sequence 553 AA;

Query Match 86.8%; Score 2470; DB 21; Length 553;
Best Local Similarity 84.6%; Pred. No. 2.3e-243;
Matches 468; Conservative 39; Mismatches 44; Indels 2; Gaps 1;

Qy 1 MRDYDEVIAFLGEMGFQRLIFPLLSASTIIPNGFNGMSVVFVLAGTPEHRCRVPDAANLSS 60
Db 1 mrdydeviaflgwgpfqrliflllsasliipngfngmsvfvfagtpchrcrlvpdtvniss 60

Qy 61 AWRNNSVPLRLRGREVPVHSCSRYLATIANFSALGLEPGRDVLQGLQESCLDGWEFS 120
Db 61 swrnhslpletkgvqpqscrrylatiansamglepggqvdleqlqesclgdgweyd 120

Qy 121 QDYLSTVVTENNYCEDNWKVPLTSLFFVYGLGVSFGSLSDRFGRKKNWLEATMAVQ 180
Db 121 kdlfstivtenwlvnceddwtktpltsffvgvlcgsvs9qldorfgrkknvlfatmaq 180
Qy 181 TGFSLQLOFSISWEMFTVLVIVGMGQISNYVAFILGTEILGKSVRIIFSTGLGVCTFFA 240
Db 181 tgfsvqifstnwemftvlfaivgmqgisnyvvaifilgteilsksvriifstlglvctffa 240
Qy 241 VGYMLPLPAYFIRDRWMLLLALTVPGLVCLVPLWFWFIPESPRWLISQRRFEAEIIOKA 300
Db 241 igymvlpfayfirdwrmlallaltlpgfcvplwfwfipesprwlisqrfaeaeqlqka 300
Qy 301 AKMNTAVPAVLDSVE--ELNPLKQOKAFILDLFRTNRIATIMTSLLLMLTSLVGVYFA 358
Db 301 akmsvavpagifdplelqelnslkqgkvilldlfrtniatitvmavmlwmltsvgvfa 360
Qy 359 LSLDAPNLHGDAYLNCFLSALIEIPAYITAWLLKRLTPRYIIAAVFLWGGGVLLFIQIV 418
Db 361 lslvnpnlhgdvylncflslgievpayftawllrtlpriiagvlfwgggvllliqv 420
Qy 419 PDYFELSLGLVWLKFGITSAFSLMYVFTAEIPLTVRNMAVGVTSTASRVGSIAPVF 478
Db 421 pedyfvsiglvmlgkfgitsafslmyvftaelvptlvrnmavgitmsarvgsiapyf 480
Qy 479 VYLGAYNRMPLPYIVMGSLTVLIGIFTLFFPESLGMPLPETLEOMQVKWFRSCKKTRDSM 538
Db 481 vylgaynrllpyilmslvtlglitlffesfgvtlpenleqmkgvrgcggkktvsv 540
Qy 539 ETEENPKVLITAF 551
Db 541 dreespkvlitaf 553

RESULT 4

AAY01650
ID AAY01650 standard; Protein; 557 AA.

XX AC AAY01650;

XX DT 23-JUN-1999 (first entry)

XX DE A protein with cation transporting activity.

XX KW Organic cation transporter; OCT1; OCT2; drug development; fatty liver;
XX KW heart disease; cancer; anti-tumour drug; anticancer drug.

XX OS Homo sapiens.

XX PN WO9913072-A1.

XX PD 18-MAR-1999.

XX PF 07-SEP-1998; 98WO-JP04009.

XX PR 20-MAY-1998; 98JP-0156660.

XX PR 08-SEP-1997; 97JP-0260972.

XX (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.

XX PI Nezu J, Oku A;

XX WPI; 1999-215062/18.

XX N-PSDB; AAX26880.

PT Genes homologous with organic cation transporters OCT1 and OCT2,
PT useful in design of new drugs for treatment of diseases due to
PT abnormality of the transporter functions

PS Claim 1; Page 51-55; 97pp; Japanese.

XX The present sequence represents a protein with cation transporting
XX activity. The genes are significantly homologous with organic cation
CC

CC transporters OCT1 and OCT2. The genes may be used in drug development,
 CC particularly in the treatment of diseases due to abnormality of the
 CC organic cation transporter functions e.g. fatty liver, heart diseases
 CC and cancers, by controlling such as by inhibition or activation.
 CC Administration of anti-tumour and anticancer drugs in combination with
 CC a transporter protein inhibiting agent allows the agents to penetrate
 CC into the diseased cells to enhance the drug action.
 XX
 SQ Sequence 557 AA;

Query Match 78.0%; Score 2218; DB 20; Length 557;
 Best Local Similarity 75.9%; Pred. No. 1.2e-217;
 Matches 423; Conservative 57; Mismatches 71; Indels 6; Gaps 3;

Qy 1 MRDDEVIAFLGEGPQRIIFLLSASIPNGFNQMSVVFELAGTPEHRCRVPDAANLSS 60
 Db 1 mrdydevtafigewgqgrliffllsasiipngftlssvfliaatpehrcrvpdaanlss 60
 Qy 61 AWRNNSVPLRLDRGVPHSCSRYRLATIANFSALGLEPGRDVDLQLEQESCLDGEWFS 120
 Db 61 awrnhvplrlrdgrevphscrryrlatiansalsglepgrdvdlqleqescldgweifs 120
 Qy 121 QDVYLSVTWTEWNLVCDNKKVPLTTSLEFVGVLLGSFVSGQLSDRFGRKNVLFATMAVQ 180
 Db 121 qdvylstvtwtewnlvcdndkvpklttslefvgvllgslfsgqlsdrgrknvlfatmgmq 180
 Qy 181 TGFSEFLQIFSIWEMFTVLFVIVGMQISNYVAVFVLGTEILGKSVRIIFSLGVCTEFA 240
 Db 181 tgfslqlfisknfemfvlvlgmgqisnyvaafvlgteilgksvriifstlglvcifya 240
 Qy 241 VGYMLLPFLFAYFIRDMRLMLLALTVPGLVPLWTFIPESPRWLISQRFREAEDIIOKA 300
 Db 241 fgymlplfayfirdmrlmlaltvpglvplwtfipspwrlisqrgfreaediioka 300
 Qy 301 AKMNTAVPAVIFD--SVEELNPKQOKAFILDFTRTNIAIMTSLMLLMLTSGVYFA 358
 Db 301 akangivvpstifdpseqlqslskkqshnildlrltwnlrmvmtslmlwtisvgyfg 360
 Qy 359 LSLDAPNLHGDAYLNCFLSALIEIPAYITAWLLRLTPRYIAAFLPWGGVLLFIQIVL 418
 Db 361 lsldapnlhgdaylncflsalieipayitawllrltpryiaaflpwggvllfioqlv 420
 Qy 419 PVDYFSLIGLVMLGKFGITSFASMLYVFTAEYPTLYRNMAVGTSTASRVGSIAPYF 478
 Db 421 pddiylatvlvmgkfgvtaafsmvvytaelyptwvnmvgvstasrslgsilspvf 480
 Qy 479 VYLGAYNRMLPYIVMGLTIVLIGITLFFPESLGMTLPETLEOMQKVKWFRSGK---KTR 535
 Db 481 vylgaydrfplylmgsltiltaltlfpesfgtpltidqmlrvkmkhrktpshtr 540
 Qy 536 DSMETENPKVL-ITAF 551
 Db 541 mlkdqgerptilkstaf 557

RESULT 5
 AA83929
 ID AA83929 standard; Protein: 557 AA.
 XX
 AC AA83929;
 XX
 DT 05-JUL-2000 (first entry)
 XX
 DE Human carnitine transporter protein OCTN2.
 KW Organic cation transportation; human; carnitine transporter protein;
 KW OCTN2; diagnosis; systemic carnitine deficiency; mutation; gene therapy;
 KW juvenile visceral steatosis.
 XX
 OS Homo sapiens.
 XX
 PW W0200014210-A1.

XX 16-MAR-2000.
 PD
 PF 07-SEP-1999; 99WO-JP04853.
 XX
 XX 07-SEP-1998; 98JP-0252683.
 PR
 XX (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
 PA
 XX Nezu J, Oku A;
 PI
 XX WPI; 2000-256966/22.
 DR
 DR N-PSDB; AAA09889.
 XX
 PT Systemic carnitine deficiency gene OCTN2 encoding part of organic
 PT cation transporter, useful as diagnostic tool -
 PS Example 1; Page 36-40; 106pp; Japanese.
 XX
 CC This sequence represents the human carnitine transporter protein OCTN2.
 CC The coding sequence can be used as a target for diagnosis of systemic
 CC carnitine deficiency by detecting the presence of mutations in the
 CC sequence, especially seen in the disease juvenile visceral steatosis
 CC (jvs). The wild type OCTN2 gene can be used in the gene therapy of the
 CC disease state.
 CC
 XX Sequence 557 AA;

Query Match 78.0%; Score 2218; DB 21; Length 557;
 Best Local Similarity 75.9%; Pred. No. 1.2e-217;
 Matches 423; Conservative 57; Mismatches 71; Indels 6; Gaps 3;

Qy 1 MRDDEVIAFLGEGPQRIIFLLSASIPNGFNQMSVVFELAGTPEHRCRVPDAANLSS 60
 Db 1 mrdydevtafigewgqgrliffllsasiipngftlssvfliaatpehrcrvpdaanlss 60
 Qy 61 AWRNNSVPLRLDRGVPHSCSRYRLATIANFSALGLEPGRDVDLQLEQESCLDGEWFS 120
 Db 61 awrnhvplrlrdgrevphscrryrlatiansalsglepgrdvdlqleqescldgweifs 120
 Qy 121 QDVYLSVTWTEWNLVCDNKKVPLTTSLEFVGVLLGSFVSGQLSDRFGRKNVLFATMAVQ 180
 Db 121 qdvylstvtwtewnlvcdndkvpklttslefvgvllgslfsgqlsdrgrknvlfatmgmq 180
 Qy 181 TGFSEFLQIFSIWEMFTVLFVIVGMQISNYVAVFVLGTEILGKSVRIIFSLGVCTEFA 240
 Db 181 tgfslqlfisknfemfvlvlgmgqisnyvaafvlgteilgksvriifstlglvcifya 240
 Qy 241 VGYMLLPFLFAYFIRDMRLMLLALTVPGLVPLWTFIPESPRWLISQRFREAEDIIOKA 300
 Db 241 fgymlplfayfirdmrlmlaltvpglvplwtfipspwrlisqrgfreaediioka 300
 Qy 301 AKMNTAVPAVIFD--SVEELNPKQOKAFILDFTRTNIAIMTSLMLLMLTSGVYFA 358
 Db 301 akangivvpstifdpseqlqslskkqshnildlrltwnlrmvmtslmlwtisvgyfg 360
 Qy 359 LSLDAPNLHGDAYLNCFLSALIEIPAYITAWLLRLTPRYIAAFLPWGGVLLFIQIVL 418
 Db 361 lsldapnlhgdaylncflsalieipayitawllrltpryiaaflpwggvllfioqlv 420
 Qy 419 PVDYFSLIGLVMLGKFGITSFASMLYVFTAEYPTLYRNMAVGTSTASRVGSIAPYF 478
 Db 421 pddiylatvlvmgkfgvtaafsmvvytaelyptwvnmvgvstasrslgsilspvf 480
 Qy 479 VYLGAYNRMLPYIVMGLTIVLIGITLFFPESLGMTLPETLEOMQKVKWFRSGK---KTR 535
 Db 481 vylgaydrfplylmgsltiltaltlfpesfgtpltidqmlrvkmkhrktpshtr 540
 Qy 536 DSMETENPKVL-ITAF 551
 Db 541 mlkdqgerptilkstaf 557


```

RESULT 6
AAY01652
ID AAY01652 standard; Protein; 557 AA.
XX
AC AAY01652:
XX
DT 23-JUN-1999 (first entry)
XX
DE A protein with cation transporting activity.
XX
KW Organic cation transporter; OCT1; OCT2; drug development; fatty liver;
KW heart disease; cancer; anti-tumour drug; anticancer drug.
XX
OS Mus musculus.
XX
PN WO9913072-A1.
XX
PD 18-MAR-1999.
XX
PF 07-SEP-1998; 98WO-JP04009.
XX
PR 20-MAY-1998; 98JP-0156660.
XX
PR 08-SEP-1997; 97JP-0260972.
XX
PA (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
XX
PI Nezu J, Oku A;
XX
DR WPI; 1999-215062/18.
XX
DR N-PSDB; AAX26902.
XX
PT Genes homologous with organic cation transporters OCT1 and OCT2,
PT useful in design of new drugs for treatment of diseases due to
PT abnormality of the transporter functions
XX
PS Claim 1; Page 75-79; 97pp; Japanese.
XX
CC The present sequence represents a protein with cation transporting
CC activity. The genes are significantly homologous with organic cation
CC transporters OCT1 and OCT2. The genes may be used in drug development,
CC particularly in the treatment of diseases due to abnormality of the
CC organic cation transporter functions e.g. fatty liver, heart diseases
CC and cancers, by controlling such as by inhibition or activation.
CC Administration of anti-tumour and anticancer drugs in combination with
CC a transporter protein inhibiting agent allows the agents to penetrate
CC into the diseased cells to enhance the drug action.
XX
SQ Sequence 557 AA;

Query Match 76.1%; Score 2164; DB 20; Length 557;
Best Local Similarity 72.9%; Pred. No. 3.9e-212;
Matches 406; Conservative 70; Mismatches 75; Indels 6; Gaps 3;

Qy 1 MRDYDEVIAFLGEGWGFQRLIFLLSASIIIPNGFNGMSVFLAGTPEHRKRVDPDANLSS 60
Db - 1 mrdydeviatlgewgpfqrlifllsasiipngfngmsivflagtpehrclvphntvnlss 60
Qy 61 AWRNNSVPLRLRDGVPVHSCSRRLATIANFSALEPGRDVDLGOEQESCLDGWEFS 120
Db 61 awrnhsipletkgqvqpkrrylatianfseiglepgdrvdledqesclgdgweyd 120
Qy 121 QDVLSTVVTENLVCDNKKVPLTSLFFVGLGVSGVSGVLSQSRFGKKNVLFATMAVQ 180
Db 121 kdvlstvtewdlvckddwkaptltslffvglmgsflsgqlsdrfgrknvlfatmgmq 180
Qy 181 TGFSLQIFSTISMEFTVFLVVGMSQISNYVAVFLTEILGKSVRIFFSTLGCWCFEA 240
Db 181 tgfslqfvsnvfmftvflvvgmgqlsnyvaarvlgteillsksrllfatlgvcvlfya 240
Qy 241 VGYMLLFLFAFYFIRDMRMLLALTVPGLCVPLWMPFESPRWLISQRRFREADIIOKA 300
Db 241 vgmllflfayfirdmrmlalltvpvglcgalwmpfespwlisqrrfkeaeavilrka 300

```

```

Db 241 fgfmvplfayfirdwrmlallaltvpgvlcgalwmpfespwlisqrrfkeaeavilrka 300
Qy 301 AKMNTAVPAVIED--SVEELNPLKOOKAFILDFETRNIAITMTMSLLWMLTSVGYFA 358
Db 301 akingivapstifopseiqdlnstkpqlhhiydlirtnirvitmsilwltisvgyfg 360
Qy 359 LSLDAPNLHGDAYLNCFLSALIEIPAYITAWLLRLTPRRYITAAVLFWGGGVLLFIOLV 418
Db 361 lsldtphlghdiynvcflaavepayvlawlllqylprysisaalfggsvllfmqlv 420
Qy 419 PVDYFYLISGLVMLGKGITSAFESMLYVFTAEIYPTLVNMAVGVTSTASRVGSIIAPYF 478
Db 421 pselfyistalvmvgkfgitsaysmvvytaelyptcvvrnmvgvssstasrlgsilspyf 480
Qy 479 VYLGAYNRMLPYIVMGSLTVLIGLITLFFPESLGMTLPETLEQMKQVKWFRSGK---KTR 535
Db 481 vylgaydrfplylmgsltiltaltlffpessfgvplpdtidqmlrvkgikwqdiqgstr 540
Qy 536 DSMETEENPKVL-ITAF 551
Db 541 mqkdgessptvilkstaf 557

RESULT 7
AAB20580
ID AAB20580 standard; Protein; 557 AA.
XX
AC AAB20580;
XX
DT 11-DEC-2000 (first entry)
XX
DE Mouse OCTN2 amino acid sequence.
XX
KW Mouse; transporter; OCTN1; OCTN2; OCTN3; organic cation transporter;
KW identification; regulator; carnitine transport.
XX
OS Mus musculus.
XX
PN WO2000046368-A1.
XX
PD 10-AUG-2000.
XX
PF 04-FEB-2000; 2000WO-JP00619.
XX
PR 05-FEB-1999; 99JP-0028406.
XX
PA (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
XX
PI Nezu J, Ose A;
XX
DR WPI; 2000-586982/55.
XX
PT Organic cation transporter gene OCTN3 expressed in testis for
PT identification of regulators of carnitine transport for use as drugs
XX
PS Example 2; Fig 2; 58pp; Japanese.
XX
CC The present invention describes a mouse organic cation transporter
CC protein (OCTN3). Also described are: (1) a method for screening
CC compounds for their ability to regulate the transport of an organic
CC cation into the cell, by generating a cell expressing OCTN3 at the
CC cell membrane, contacting with the compound and observing OCTN3 at the
CC observing the degree of transport of the compound and organic cation, and
CC method for screening compounds for their ability to be transported into
CC the cell by OCTN3, by generating a cell expressing OCTN3 at the cell
CC membrane, contacting with the compound and observing the degree of
CC transport of the compound. OCTN3 can be used for the identification of
CC regulators of the transport of organic cations (especially carnitine)
CC into cells by OCTN3, for use as drugs. The present sequence represents
CC the mouse OCTN2 amino acid sequence, which is used in an example from
XX the present invention.
XX Sequence 557 AA;

```

Query Match	76.1%; Score 2164; DB 21; Length 557;
Best Local Similarity	72.9%; Pred. No. 3.9e-212;
Matches 406; Conservative	70; Mismatches 75; Indels 6; Gaps 3;
QY 1	MRDYDEVIAFLGEMGPFQRLIFLLSASIIIPNGFNMGMSVVFVLAGTPEHRCRVPDANLSS 60
DB 1	mrdydevtaflegwpgfqrllifllsasiiipngfnmgmsivflagtpehrcrlvphvtnlss 60
QY 61	AWRNNSVPLRLRDGREVPHSCSYRLATIANFSALGLEPGRDVDLQLEQESCLDGEWFFS 120
DB 61	awrnhsipletkdgrqvpqkrryrlatianfselepggrdvleleqesclgdgweyd 120
QY 121	QDVYLYSTVVTWNLVCDNWKVPLTTSFLFVGVLLGSFVSGQLSDRFGRKKNVLFATMAVQ 180
DB 121	kdvflstivtewdlvckdkwkapltslffvglmgsfsglsdrfgrkknvlfatmgmq 180
QY 181	TGFSFLQIFSISWEMFTVLFVIVMGQISNYVAVFILGTEILGKSVRIIFSTGLVCTFEFA 240
DB 181	tgfsflqfsvfnfemftvlfvlgmgqisnyvaafvlgteilsksirifafatlgvcifya 240
QY 241	VCYMLLPFAFYFIRDWRMLLALTVPGLVCPVLPWWTIPESPRWLISQRRFREAEDIIOKA 300
DB 241	fgfmvlpfayfirdwrmlaltvpvlgcalwfpesprwlisqrrfreakaeavliirka 300
QY 301	AKMNTAVPAVIFD--SVEELNPLKQKAFILDLFRTNIAIMTMSLLMLTSGVGYFA 358
DB 301	akingivapstfipdselqdnstkpqlhhydltrnirvitmsilwitisvgyfg 360
QY 359	LSLDAPNLHGDAYLNCFLSALIEIPAYITAWLLRLTPRYIIAAVLFWGGVLLFIQIV 418
DB 361	lsldtphnlgdyvncfllaavepayvawllilqylpryisaaflggsavllfmqlv 420
QY 419	PVDYVFLSIGLVMLGKFGITSAFSLMVFYTAELPTLVLRNMAVGTSTASRVGSIAPYF 478
DB 421	pselfylstailvmvgkfgitsaysmvvytaelyptvvrnmvgvssstasrlgsilspyf 480
QY 479	VYLGAYNRLPYIVMGSLTVLIGITFLFFPESLGMPLPETLEOMOKVWFRSGK---KTR 535
DB 481	vylygaydrflpylmgsltiltaltlffpesfgvplpdtidqmrivkgikwqqlgsqtr 540
QY 536	DSMETEENPKVL-ITAF 551
DB 541	mqkdgdesptvltkstaf 557
RESULT 8	
AAV83930	
ID	AAV83930 standard; protein; 557 AA.
XX	
AC	AAV83930;
XX	
DT	05-JUL-2000 (first entry)
XX	
DE	Mouse carnitine transporter protein OCTN2.
XX	
KW	Organic cation transportation; mouse; carnitine transporter protein;
KW	OCTN2; diagnosis; systemic carnitine deficiency; mutation; gene therapy;
KW	juvenile visceral steatosis.
XX	
OS	Mus musculus.
XX	
PN	W0200014210-A1.
XX	
PD	16-MAR-2000.
XX	
PF	07-SEP-1999; 99W0-JF04853.
XX	
PR	07-SEP-1998; 98JP-0252683.
XX	
PA	(CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
XX	

PI	Nezu J, Oku A;
XX	
DR	WPI: 2000-256966/22.
DR	N-PSDB; AAA09890.
XX	
PT	Systemic carnitine deficiency gene OCTN2 encoding part of organic
PT	cation transporter, useful as diagnostic tool
XX	
PS	Example 1; Page 47-51; 106pp; Japanese.
XX	
CC	This sequence represents the mouse carnitine transporter protein OCTN2.
CC	The coding sequence of the corresponding human protein can be used as a
CC	target for diagnosis of systemic carnitine deficiency by detecting the
CC	presence of mutations in the sequence, especially seen in the disease
CC	juvenile visceral steatosis (jvs). The wild type OCTN2 gene can be used
CC	in the gene therapy of the disease state.
XX	
SQ	Sequence 557 AA;
	Query Match 76.1%; Score 2164; DB 21; Length 557;
	Best Local Similarity 72.9%; Pred. No. 3.9e-212;
	Matches 406; Conservative 70; Mismatches 75; Indels 6; Gaps 3;
QY 1	MRDYDEVIAFLGEMGPFQRLIFLLSASIIIPNGFNMGMSVVFVLAGTPEHRCRVPDANLSS 60
DB 1	mrdydevtaflegwpgfqrllifllsasiiipngfnmgmsivflagtpehrcrlvphvtnlss 60
QY 61	AWRNNSVPLRLRDGREVPHSCSYRLATIANFSALGLEPGRDVDLQLEQESCLDGEWFFS 120
DB 61	awrnhsipletkdgrqvpqkrryrlatianfselepggrdvleleqesclgdgweyd 120
QY 121	QDVYLYSTVVTWNLVCDNWKVPLTTSFLFVGVLLGSFVSGQLSDRFGRKKNVLFATMAVQ 180
DB 121	kdvflstivtewdlvckdkwkapltslffvglmgsfsglsdrfgrkknvlfatmgmq 180
QY 181	TGFSFLQIFSISWEMFTVLFVIVMGQISNYVAVFILGTEILGKSVRIIFSTGLVCTFEFA 240
DB 181	tgfsflqfsvfnfemftvlfvlgmgqisnyvaafvlgteilsksirifafatlgvcifya 240
QY 241	VCYMLLPFAFYFIRDWRMLLALTVPGLVCPVLPWWTIPESPRWLISQRRFREAEDIIOKA 300
DB 241	fgfmvlpfayfirdwrmlaltvpvlgcalwfpesprwlisqrrfreakaeavliirka 300
QY 301	AKMNTAVPAVIFD--SVEELNPLKQKAFILDLFRTNIAIMTMSLLMLTSGVGYFA 358
DB 301	akingivapstfipdselqdnstkpqlhhydltrnirvitmsilwitisvgyfg 360
QY 359	LSLDAPNLHGDAYLNCFLSALIEIPAYITAWLLRLTPRYIIAAVLFWGGVLLFIQIV 418
DB 361	lsldtphnlgdyvncfllaavepayvawllilqylpryisaaflggsavllfmqlv 420
QY 419	PVDYVFLSIGLVMLGKFGITSAFSLMVFYTAELPTLVLRNMAVGTSTASRVGSIAPYF 478
DB 421	pselfylstailvmvgkfgitsaysmvvytaelyptvvrnmvgvssstasrlgsilspyf 480
QY 479	VYLGAYNRLPYIVMGSLTVLIGITFLFFPESLGMPLPETLEOMOKVWFRSGK---KTR 535
DB 481	vylygaydrflpylmgsltiltaltlffpesfgvplpdtidqmrivkgikwqqlgsqtr 540
QY 536	DSMETEENPKVL-ITAF 551
DB 541	mqkdgdesptvltkstaf 557
RESULT 9	
AAB20578	
ID	AAB20578 standard; Protein; 564 AA.
XX	
AC	AAB20578;
XX	
DT	11-DEC-2000 (first entry)
XX	

DE Mouse OCTN3 protein SEQ ID NO:1.
 XX Mouse; transporter; OCTN1; OCTN2; OCTN3; organic cation transporter;
 KW identification; regulator; carnitine transport.
 XX Mus musculus.
 XX WO2000046368-A1.
 XX 10-AUG-2000.
 XX 04-FEB-2000; 2000WO-JP00619.
 XX 05-FEB-1999; 99JP-0028406.
 XX (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
 XX Nezu J, Ose A;
 XX WPI: 2000-586982/55.
 XX N-PSDB; AAA88053.
 XX Organic cation transporter gene OCTN3 expressed in testis for
 PT identification of regulators of carnitine transport for use as drugs
 PT
 XX Claim 1; Page 34-39; 58pp; Japanese.
 XX The present invention describes a mouse organic cation transporter
 CC protein (OCTN3). Also described are: (1) a method for screening
 CC compounds for their ability to regulate the transport of an organic
 CC cation into the cell, by generating a cell expressing OCTN3 at the
 CC cell membrane, contacting with the compound and organic cation, and
 CC observing the degree of transport of the organic cation; and (2) a
 CC method for screening compounds for their ability to be transported into
 CC the cell by OCTN3, by generating a cell expressing OCTN3 at the cell
 CC membrane, contacting with the compound and observing the degree of
 CC transport of the compound. OCTN3 can be used for the identification of
 CC regulators of the transport of organic cations (especially carnitine)
 CC into cells by OCTN3, for use as drugs. The present sequence represents
 CC mouse OCTN3.
 XX
 SQ Sequence 564 AA;

Query Match 71.5%; Score 2034.5; DB 21; Length 564;
 Best Local Similarity 69.0%; Pred. No. 6.5e-199;
 Matches 380; Conservative 75; Mismatches 85; Indels 11; Gaps 3;

QY 1 MRDYDEVIAFLGEMGPFORLIRFLLSASTIPNGFNGMSVVELAGTPEHRCRVPDAANLSS 60
 Db 1 mldydevtafigewgtfqlrllfllsasiipngftgisavftaipchrcripdtvniss 60
 QY 61 AWRNNSVPLRLDRGVRPHSCSRYLRIATNFSALGLEPGRVDLGLQLESCLDGWERS 120
 Db 61 awrnhsipmetkdgpevpqkrcrryrlatnfsalglepgrvdldleqleqncldgweyd 120
 QY 121 ODVYLSVTVENLVEDNWKVPLTSLFVGVLLGSFVSGQLSDRPGKKNVLFATMAVO 180
 Db 121 kdiflstivtewdlvckddkaptltsffvgyvllgfsisqglssdrfgrknllftmnh 180
 QY 181 TGFSELIQISFISWEMTFLFVIVGMGOISNYVVAFTLGTETILGKSVRIITFSLGVTCEFA 240
 Db 181 tgrsfliqvsvnfemflltylvgmghisnyvaafvlgtemisksvrllfatlgwcfifa 240
 QY 241 VGYMLPLFAFYIRDMRLMLLALTVPGLVCLPWLWFPISPRWLISQRRFREAEIIOKA 300
 Db 241 fgfmvlpfayfirewrrlllatlpgvicgalwvfipspwrlisqgrikaeviirka 300
 QY 301 AKMNTAVPAVIFDSVEELNPL-----KQKAFILDLFRTRNIAITMTMSLLMLWLTSV 354
 Db 301 akingivapstifdp-setnklqddskkpkqshhdyldvrtptnrltmsilwitiv 359
 QY 355 GYFALSIDAPNLHGDAYLNCFLSALIEIPAYITAWLLLTPLPRYIAAVLFWGGGVLLF 414

Db 360 gyfslsldtpnlnynvncflllaavevpayvialwlllghvrrysmagslfiggsvlll 419
 QY 415 IQLVPDYDYFLSIGLVMGLKFGTSAFSLMYVFTAEIYPTLVRNMAVGVSTASRVGSII 474
 Db 420 vqlvpsdlhlylstdlvmvgkfgitsaysmvyvytaelyptvvrnmvgvsvstasrlgsil 479
 QY 475 APYFVILGAYNRMLPYIVMGLTFLVIGLIFLFPFESLGMILPETLEOMQKVKWFRSGKKT 534
 Db 480 spfyvgyaydrripyllmsltliltaitlfpessgvsipetidemqkvkl-----kg 535
 QY 535 RDSMETEENPK 545
 Db 536 rqsiskkgspk 546
 RESULT 10
 AAB43038
 ID AAB43038 standard; Protein; 560 AA.
 XX
 AC AAB43038;
 XX
 DT 08-FEB-2001 (first entry)
 XX
 DE Human ORFX ORF2802 polypeptide sequence SEQ ID NO:5604.
 KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
 KW vulnary; antiprosoratic; antiparkinsonian; nootropic; neuroprotective;
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
 KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
 KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
 KW thrombosis; contraceptive.
 XX
 OS Homo sapiens.
 XX
 PN WO2000058473-A2.
 XX
 PD 05-OCT-2000.
 XX
 PF 31-MAR-2000; 2000WO-US08621.
 XX
 PR 31-MAR-1999; 99US-0127607.
 PR 02-APR-1999; 99US-0127636.
 PR 05-APR-1999; 99US-0127728.
 PR 30-MAR-2000; 2000US-0540763.
 XX
 PA (CURA-) CURAGEN CORP.
 XX
 PI Shimkets RA, Leach M;
 XX
 DR WPI; 2000-602362/57.
 XX N-PSDB; AAC77247.
 PT Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease -
 XX
 XX Claim 11; Page 4791-4792; 5507pp; English.
 CC AACH74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytostatic; hepatotropic; vulnary;
 CC antiprosoratic; antiparkinsonian; nootropic; neuroprotective;
 CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;

CC antidiabetic; hypcensive; dermatological; immunosuppressive;
 CC antiinflammatory; antibacterial; antifungal; antirheumatic;
 CC antithyroid; and antianemic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORFX-associated disorder. The
 CC nucleic acids can be used to express ORFX proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.

XX Sequence 560 AA;

Query Match 27.4%; Score 779.5; DB 21; Length 560;
 Best Local Similarity 32.9%; Pred. No. 7.7e-71;
 Matches 184; Conservative 113; Mismatches 214; Indels 49; Gaps 9;

QY 17 FORLIFLLSASITPNFGMSVFLAGTPEHRCRP-----DAANLS 59
 Db 2 fqrlylfcagfniscgihylasvfmgtvphrcpgrvnsqvfhnswnsltdtgall 61
 QY 60 SAWRNSVPLRLDGRVPHS-CSRYRLATIANFSALEGRDVLGQLESCLDGWE 118
 Db 62 ssqkdyvqlqgeiwelrcsrnkr---entsslgve-----ytgskkefcvdygi 113
 QY 119 FSQDVLSTVTVENLVCEDNWKVPLTSLFFVGLVLSFVSGQSLDRGRKNVLPATMA 178
 Db 114 ydqntwkstvtqnlvcdkrlamlqlfmgfvlgsvtgfyfdrgrvrlwatss 173
 QY 179 VQTGFSFLOISISWEMFTLVFVVGMSQISNVAVILGTEILGKSVRIFFSLGVCTP 238
 Db 174 smflglaaafavdyttfmaarfflanvasgylvgvfyvmefimgksr-twasvhlhsf 232
 QY 239 FAVGYMLPLFAYFTRDW---RMLLALTVPGLVCLVPLWFWIPESRWLISORFREAE 295
 Db 233 favgtllvalgyivrtwlyqmlstvtvpfllcc---wvlpetfllsegrveeaq 289
 QY 296 IQKAANKNNTA-----VPAVIFDSVEELNPLKQOKAFILDLFTRTRNIAIMTSL 349
 Db 290 lvdImakwnrasscklsellidlgppvsnspetevqkhnlsylfynwsitkrltvlwliw 349
 QY 350 MLTSVGFALSADPNLHGDAYLNCFLSALIEIPAYITAMLLLTLPRTVIAAVLFWGG 409
 Db 350 ftgslgfyfslsvnlggneylnflgvveipayfvciamdkvgrtvlayslfcsa 409
 QY 410 GVLLFIQVLPDYFSLSIGLVMKFGITSAFSLMYVFTAEVPTLVVRNMAVGVTSTASR 469
 Db 410 lacgvnmvlpkqkylilgvvtamvqkfaigaafglylytaelptvirsavsgsmvcr 469
 QY 470 VGSIIAPFVVLGYANRMLPYVINGMSVTVLIGITLFFPESLGMTLPETLEQKQKWKFR 529
 Db 470 lasilapfsvdlsliwifpqlfvtmallsgvltklpeltgkrlattweeaakl---- 525
 QY 530 SGKTRDSMETENPKVLIT 549
 Db 526 -----eseneskskllit 539

RESULT 11
 AAB49401
 ID AAB49401 standard; Protein; 540 AA.
 XX
 AC AAB49401;
 XX
 DT 02-MAR-2001 (first entry)
 XX
 DE Murine organic anion transporter 6.

XX Murine; organic anion transporter 6; mOATP6; cancer; inflammation;
 KW cardiovascular disease; central nervous system disorder; kidney disease;
 KW liver disease; autoimmune disease.
 XX Mus sp.
 OS WO200070048-A1.
 PN 23-NOV-2000.
 XX 15-MAY-2000; 2000WO-US13316.
 PF 14-MAY-1999; 99US-0134137.
 PR 12-MAY-2000; 2000US-0570293.
 XX (SMIK) SMITHKLINE BEECHAM CORP.
 PA (SMIK) SMITHKLINE BEECHAM PLC.
 XX Feild J, Yue L, Ellens H;
 PI WPI; 2001-016235/02.
 XX N-PSDB; AAC83979.
 DR Murine organic anion transporter 6 polypeptide, useful for identifying
 PT agonists/antagonists that are useful in treatment of cancer, kidney
 PT disease, autoimmune disease, inflammation and cardiovascular disease
 XX Claim 2: Page 28-29; 32pp; English.
 PS The present sequence is murine organic anion transporter 6 (mOATP6).
 CC mOATP6 protein is useful for screening compounds which inhibit or
 CC stimulate the function of mOATP6 and also compounds that neither agonise
 CC nor antagonise OATP6. The identified agonists and antagonists are useful
 CC for prevention and treatment of human diseases, including cancer,
 CC inflammation, cardiovascular disease, central nervous system disorders,
 CC kidney diseases, liver disease and autoimmune diseases.
 CC Sequence 540 AA;

Query Match 24.5%; Score 698; DB 22; Length 540;
 Best Local Similarity 32.5%; Pred. No. 1.5e-62;
 Matches 183; Conservative 97; Mismatches 231; Indels 52; Gaps 19;

QY 4 YDEVIATFLGEMGPRQ--RLIFFLLSASITPNFGMSVFLAGTPEHRCRPDA-ANLS- 59
 Db 3 feelhkvvggfgpqlrnlvlalprflpmhf--llpifmaavpahhcalpdpapanish 60
 QY 60 -SAWRNSVPLRLDGRVPHS-CSRYRLATIANFSALEGRDVLGQLEQ-----SCL 114
 Db 61 qdlwkthlp-retgdsf--ssclrfaypqaipnvtlgtv---ynsgepegepltvpc 114
 QY 115 DQWFSQDVLSTVTVENLVCEDNWKVPLTSLFFVGLVLSFVSGQSLDRGRKNVLF 174
 Db 115 qgweydrsefsstiatetwldvcqrglnkvtctcfilgavvyeylsdrfgrrill 174
 QY 175 ATMAVQTGFSFLOISISWEMFTLVFVVGMSQISNVAVILGTEILGKSVRIFFSLTG 234
 Db 175 vavvstlalgmsaasvnyimfvtrmtlgsalagftiivlplewldvhehrtvagvis 234
 QY 235 VCTFFAVGYMLPLFAYFTRDWRMMLLALTVPGLVCLP-----LWMFIPESRWLISQRRF 290
 Db 235 t-tftwggvllltlvgylrsrwillaatlp---cvpgilsiw-vpesarwlltqgrv 289
 QY 291 REAEDIQKAANKNNTAVPAVIFDSV--EELNPL-----KQOKAFILDLFTRTRNIAIMI 343
 Db 290 eakkyisicakinpripse---dsisqeaalkvmtmervsqirpsylldlfrtsqrlhsvl 346
 QY 344 MSLLLWMLTSGVGFALSADPNLHGDAYLNCFLSALIEIPAYITAMLLLTLPRTVIAAV 398
 Db 347 ccmnmwfgvnsyygltidasglvtvqtlqllfgavepskitvflvrlvgrlrlteag 406

QY 399 YIIAAVLFWGGVLLFTQLVPDYVYFSLISGLVMLGRFGITSAFSMLYVFTAEIYPLTVRN 458
 Db 407 mlalaltfgisi-----lvssdtkswitalvvgkafseaafttaylftselypvlrq 461
 QY 459 MAYGVSTASRVGSIITAPYFVYLGAYNRMPLPYIVMGSLLVILGIFTLFFPESLGWTLPET 518
 Db 462 tgmftaligraslaplvllldgvmlllpklayggisflaactvlllpctkkaqlpet 521
 QY 519 LEOMOKVWFRSGKTRDSMETE 541
 Db 522 iqdve-----rkgrki-drsqte 538

RESULT 12
 AAW64538
 ID AAW64538 standard; Protein; 554 AA.
 XX
 AC AAW64538;
 XX
 DT 21-OCT-1998 (first entry)
 XX Human liver cell clone HP01293 protein.
 XX Transmembrane domain; human; nutrition; cytokine; cell proliferation;
 KW differentiation; immune system; stimulator; suppressor; regulator;
 KW hematopoiesis; activin; inhibitor; chemotactic; chemokinetic; receptor;
 KW haemostatic; thrombolytic; ligand; anti-inflammatory; tumour.
 XX Homo sapiens.
 OS
 XX WO9821328-A2.
 PN
 XX 22-MAY-1998.
 PD
 XX 07-NOV-1997; 97WO-JP04056.
 PF
 XX 13-NOV-1996; 96JP-0301429.
 PR
 XX (PROT-) PROTEGENE INC.
 PA (SAGA) SAGAMI CHEM RES CENTRE.
 XX
 PI Kato S, Kobayashi M, Sekine S, Yamaguchi T;
 XX
 DR WPI; 1998-297932/26.
 DR N-PSDB; AAV49558, AAV49559.
 XX
 PT Human protein having transmembrane domain - useful for, e.g.
 PT research and nutrition
 PS Claim 1; Page 96-98; 205pp; English.

AAW64534-W64558 represent human proteins containing a transmembrane domain. These proteins can be used for, e.g. research and nutrition, and may have cytokine and cell proliferation/differentiation, immune stimulating/suppressing, haematopoiesis regulating, tissue growth, activin/inhibin, chemotactic/chemokinetic, haemostatic and thrombolytic, receptor/ligand, anti-inflammatory or tumour inhibition activity.
 XX
 SQ Sequence 554 AA;

Query Match 24.1%; Score 685.5; DB 19; Length 554;
 Best Local Similarity 33.0%; Pred. No. 3e-61;
 Matches 190; Conservative 87; Mismatches 229; Indels 69; Gaps 14;

QY 1 MRDYDEVIAFLGWGPFQR---LIFLLSASIIIPNGFMGSVVFLAGTPEHRCRVPDAAN 57
 Db 1 mptvddileqvgesgfwkqafllcllcllsaaafp---icvgivfigftpdhchcspgva 57
 QY 58 LSS--AWR-----NNSVPLRLDRDREVPVHSCRYR-----LATIA-NFSAIG 96
 Db 58 lsqrcwspaeelnvtpvqlpageaflgqcrreyevdwngsalscvdpplasiatnrshlp 117

QY 97 LEPRDVLGLOEQESCLDGHFEFSQDYVLSVTVTWNLVCEDNWKPVLTTLSLFFVGVLLG 156
 Db 118 lgp-----ccdgwvy--dtpgssivtefnlvccadswkldlfdscnagffg 162
 QY 157 SFVSQOLSDRGRKNVLFATMAVQTGFSLQIFISISWEMFTVLEFVIVGMGQISNYVVAFI 216
 Db 163 slvgvyfadrgrklcllgtvlvnavsgvlnafspnymmlflrlgqlvskgnwmagyt 222
 QY 217 LGTEILG----KSVRIIFSTLGVCTFFFAVGVMMLPLFAFYFDRWRMLLALATVPGVLCVP 272
 Db 223 litefvgsgsrrtvalmygma-----ftvgivaigtlayalphwrwqlavslptfll 277
 QY 273 LWWFIPESPRWLISORRFEAEIIIOKAKMNNTAVPAVI-----FDSVEELNPLKQOK 326
 Db 278 ywcvpesprwllsqnrnteaiklmhdiaqngklppadikmlsleedvteklsp----- 332
 QY 327 AFILDLFTRNIAIMTMSLLMLTSSVGYFALSDDANLHGDAYLNCFLSALIEIPAYI 386
 Db 333 -sfadlfrtprlkrftfilylwfstdsvlyggllhmgatsgnyldflysalveipdaf 391
 QY 387 TAWLLRLTPRRIITAAVLFWGGVLLFTQLVPDYVYFSLISGLVMLGRFGITSAFSMLYV 446
 Db 392 ialitidrvgrlypmavsnllagaaclvmifispdhlwlnliimcvgzmgtiaiqmcl 451
 QY 447 FTAELYPTLVRNMAVGVSTASRVGSIITAPYFVY-IGAYNRMPLPYIVMGSITVLIGIFTL 505
 Db 452 vnaelyptfvrnlgmvcslcdigiltpfivfrlrevwqalplllfavlgllaagvtl 511
 QY 506 FPEPSGLMTLPETLEOMQKVWFRSGKTRDSMET 540
 Db 512 llpetkgvalpetmkaenl-----grkakpkent 541

RESULT 13
 AAY51249
 ID AAY51249 standard; Protein; 535 AA.
 XX
 AC AAY51249;
 XX
 DT 14-APR-2000 (first entry)
 XX Rat liver anion transporter protein OAT2.
 DE
 XX OAT2; anion transporter; liver; rat; dicarboxylic acid; prostaglandin;
 KW non-steroid anti-inflammatory agent; anti-tumor.
 XX
 OS Rattus sp.
 XX JP11346779-A.
 PN
 XX 21-DEC-1999.
 PD
 XX 03-JUN-1998; 98JP-0169174.
 PF
 XX 03-JUN-1998; 98JP-0169174.
 PR
 XX (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.
 PA
 XX WPI; 2000-109694/10.
 DR N-PSDB; AAZ44679.
 XX
 XX A liver-specific org. anion transporter - can transport various
 PT substances including dicarboxylic acids
 PT
 XX Claim 2; Page 5-6; 13pp; Japanese.

XX This invention describes a novel rat liver-specific org. anion
 CC transporter OAT2. The liver-specific org. anion transporter OAT2
 CC can transport various substances including dicarboxylic acids,
 CC prostaglandins, non-steroid anti-inflammatory agents and anti-tumor
 CC agents. This sequence represents the rat liver anion transporter OAT2
 CC protein described in the method of the invention.
 XX

FT		/note= "putative N-glycosylation site"	
FT	Domain	154..171	
FT		/note= "transmembrane domain"	
FT	Domain	178..197	
FT		/note= "transmembrane domain"	
FT	Domain	243..260	
FT		/note= "transmembrane domain"	
FT	Domain	267..283	
FT		/note= "transmembrane domain"	
FT	Domain	350..366	
FT		/note= "transmembrane domain"	
FT	Domain	380..398	
FT		/note= "transmembrane domain"	
FT	Domain	406..425	
FT		/note= "transmembrane domain"	
FT	Modified-site	432	
FT		/note= "putative N-glycosylation site"	
FT	Domain	435..452	
FT		/note= "transmembrane domain"	
FT	Domain	469..485	
FT		/note= "transmembrane domain"	
FT	Domain	494..514	
FT		/note= "transmembrane domain"	
XX	DB4242577-Al.		
XX			
XX	18-JAN-1996.		
XX			
XX	13-JUL-1994;	94DE-4424577.	
XX			
XX	13-JUL-1994;	94DE-4424577.	
PR	(FARH) HOECHST AG.		
PA	Gorboulev V, Gruendeman D, Koepsell H;		
PI			
XX	WPI; 1996-069422/08.		
DR	N-PSDB; AAT08702.		
DR			
XX			
PT	Transporter protein for cationic xenobiotic(s) and pharmaceuticals,		
PT	and related DNA and transformed cells - used e.g. to assess		
PT	excretion and resorption of cationic cpds.		
XX			
PS	Claim 1; Fig 2A1; 13pp; German.		
CC			
XX	This is the amino acid sequence of the rat OCT-1 protein, a new		
CC	transporter protein able to transport cationic xenobiotics and		
CC	pharmaceuticals from the blood into liver or kidney epithelial		
CC	cells or from the intestine. The gene was isolated by injecting		
CC	a rat kidney gene library into Xenopus laevis oocyte and isolating		
CC	clones conferring uptake of 14C-tetraethylnammonium. One clone		
CC	designated OCT-1 was isolated. Expression of the gene was detect		
CC	in proximal renal tubule cells, in liver epithelial cell and in		
CC	intestinal enterocytes. The DNA can be used to generate transgenic		
CC	cells for use in vitro test for renal/biliary excretion or		
CC	intestinal resorption of xenobiotics and pharmaceuticals. The protein		
CC	of cells expressing it can also be used to isolate modulators that		
CC	block uptake of pharmaceutical by the renal tubules.		
XX			
SQ	Sequence 556 AA;		
	Query Match 24.0%; Score 681.5; DB 17; Length 556;		
	Best Local Similarity 33.0%; Pred. No. 7.6e-61;		
	Matches 183; Conservative 88; Mismatches 231; Indels 53; Gaps		
QY	1 MRDYDEVIAFGPGPRLIFELL---SASIIIPNGFMGSVVFAGTPEHRCRVPDAA 57		
	: : : : : : : : : :		
Db	1 mptcdvdlteqvgfgrgkqgflllclislap----iygviflgfcpgycnpyvae 57		
	: : : : : : : : : :		
QY	58 LSS--AWR-----NNSPV-LRLRDGREVPHSCSRYL-----ATIANFSAL 95		
	: : : : : : : : : :		
Db	58 lscrqwscaeeelnvtvqqlqpsdeasflsgcmrvegdwngstldcvdpilsslvanssq 117		
	: : : : : : : : : :		

QY 96 GLEPGRDVLDGQLEQESCLDGEFSDQVYLSVTVTWNLVCEENWVPLTTLTFFVGVLL 155
 Db 118 plp-----cehgwy--dtpgsivtefnlvcgdawkvdlfgscvnlgffl 162
 QY 156 GSFVQSLSRDRGRKNVLPATMAVOTGFSFIQFISISWEMTVLFIIVGQISNYVAF 215
 Db 163 gslvgyiadrfgrkclivtlvtlsvsgvitavapdytsmlfrllqgmvskgwvsy 222
 QY 216 ILGTEILGKSVR----IIFSTLGCVTFEAVGVMLPLFAYFIRDRWMLLALVPGVLCV 271
 Db 223 tlitefvsgyrrttailqyma-----ftvglvlgavayaipdrwqlavsiptflfi 277
 QY 272 PLMWFIPEPRWLIISORPREAEDIIOKAAMNNTAVPAVIFDSVEELNPLKOOKAFILD 331
 Db 278 lywfyvpsrwlsgkrttravimeqlagknkvppadikmicleedasekrspsfad 337
 QY 332 LFRTRNIAITMISLLMLTSGVGFALSPLDAPNLHGDAYLNCFLSALIEIPAYITAWLL 391
 Db 338 lfrtptlrkhtvilmvlfscavlygglmhvgtaganlyldfysvivefpaailvt 397
 QY 392 LRTLPRIYIAAVLFWGGVLLFIQLVPDYVYFISIGLVMIGKFGITSAFSMLYVETAEL 451
 Db 398 idrigtlyplasnltgeacllmifiphelhwnvtlclgrmgativlqmvclvnael 457
 QY 452 YPTLVNRNAVGTSTASRVGSIAPFYV-LGAYNRMLPYIVMGSLVLTIGITLFPES 510
 Db 458 yptfirnlgmvcalscdigiftbmfvrimevwaqalpilfgvlgitagamtlllpet 517
 QY 511 LGMTLPETLFOQKV 525
 Db 518 kgalpetieaeai 532
 RESULT 15
 AAB08823
 ID AAB08823 standard; Protein; 548 AA.
 AC AAB08823;
 XX 02-JAN-2001 (first entry)
 DE A human organic anion transporter 6 (OATP6).
 KW Human; organic anion transporter 6; OATP6; cancer; inflammation;
 KW cardiovascular disease; central nervous system disorder;
 KW autoimmune disease; kidney disease; liver disease.
 XX Homo sapiens.
 OS WO200052024-A1.
 PN 08-SEP-2000.
 XX 28-FEB-2000; 2000WO-US04952.
 PF 04-MAR-1999; 99US-0262563.
 PR (SMIK) SMITHKLINE BEECHAM CORP.
 PA Ellens H, Field J, Yue L;
 PI WPI: 2000-579270/54.
 DR N-PSDB; AAA75034.
 XX OATP6 polypeptides and polynucleotides encoding them, useful for
 PT treating cancer, inflammation, cardiovascular disease, central nervous
 PT system disorders, autoimmune disease, kidney or liver disease
 XX Claim 1; Page 31-32; 37pp; English.
 XX The present sequence represents human organic anion transporter 6 (OATP6)
 CC polypeptide. The OATP6 polypeptides and polynucleotides are useful for

CC treating cancer, inflammation, cardiovascular disease, central nervous
 CC system disorders, autoimmune disease, kidney and liver disease, or in
 CC identifying individuals who may handle drugs differently than normal
 CC individuals. They may also be used in identifying agonists and
 CC antagonists/inhibitors, which may be used in treating conditions
 CC associated with OATP6 imbalance, and to configure screening methods for
 CC detecting the effect of added compounds on the production of mRNA and
 CC polypeptides in cells. The polypeptides may be further used to
 CC establish assays to predict oral absorption and pharmacokinetics, and
 CC thus enhance compound and formulation design, as immunogens to produce
 CC immunospecific antibodies, and in identifying soluble or membrane-bound
 CC receptors. The polynucleotides may also be used as hybridisation probes
 CC or as primers for nucleic acid amplification, as diagnostic reagents,
 CC which may aid in the diagnosis of or susceptibility to a disease, and
 CC for chromosome identification.
 XX
 SQ Sequence 548 AA;
 Query Match 23.0%; Score 653.5; DB 21; Length 548;
 Best Local Similarity 32.0%; Pred. No. 5.4e-58;
 Matches 180; Conservative 88; Mismatches 250; Indels 45; Gaps 18;
 QY 4 YDEVIAFLGEGWGPQ--RLIFFELLSAIIIPNGFNGSVVFLAGTEPHERCRVPDA-ANLS- 59
 Db 3 feeileqvgvgfqlrnvallalprvllphf--llpflaavpahrcalpgpanfsh 60
 QY 60 -SAWRNNSVPLRDRGREGVPHSCSYRLATIANFSAIGLEPGRDVDLQLEQE-----SCL 114
 Db 61 qdvwleahlp-repdg--tlssclrfayqpaltntllgee---rqsrgeledepaycps 114
 QY 115 DGWEFSQDVYLSVTVIE--WNLVCEDNWKVPLTSLFFVGVLLGSFVSGQSLDRGRKNV 172
 Db 115 ggweydhsefstlasesqdwlvceqgnraastffagvlvgavagysdrfgrrrl 174
 QY 173 LFATMAVQTGFSFLQIFSIWEMFTVLFVIVGMGQISNVYVAFILGTETLGKSVRIIFST 232
 Db 175 llvayvstlvlgasaasvymfaiirtltgsalagftlivmplelewidvehrtvagv 234
 QY 233 LGVCTTFVAVGYMLLPFAYFIRDRWMLLALVPGVLCV-----LWTFIPSPRLISQR 288
 Db 235 ls-stftwtggvmlalvgvliirdwrllavtlp---capgillsiww-vpesarwiltqg 289
 QY 289 RFREAEDIIQKAAMNNTAVPAVIFDSVEELNPLKOOKAFI-----LDLFRTRNIAITMI 343
 Db 290 hvkeahryllhcarlngprvcedsf-sgeavskvaagervrpsyldftrprlrhisl 348
 QY 344 MSLLWMLTSGVGFALSPLDAPNLHGDAYLNCFLSALIEIPAYITAWLLRLPRYITAA 403
 Db 349 ccvvwvfgvfyyslsvdsglgnvyqtlfgavelpskllvylsvryagriltqag 408
 QY 404 VLFWGGGVLLF--IQLVVPDYVYFISIGLVMIGKFGITSAFSMLYVETAELPTLVNRNAV 461
 Db 409 tll--gtalafgrtrllvssdmkswtvavmgkafseaafttaylftseelyptvrlrqtgm 466
 QY 462 GVTASTARVGSIIAPYFVYLGAYNRMLPYIVMGSLVLTIGITLFPESLGMTLPETLEQ 521
 Db 467 gitalvgrlgslaplaallldgvmwlsipkityggiallaagtalilpctrqaipetiqd 526
 QY 522 MOKVKWFRSGKKTDRDSMETEENP 544
 Db 527 ver-----ksaptslqeemp 542

Search completed: August 16, 2001, 13:58:25
 Job time: 207 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 16, 2001, 13:56:53 ; Search time 28.45 Seconds

(without alignments)

1475.296 Million cell updates/sec

Title: US-09-521-195-1

Perfect score: 2845

Sequence: 1 MRDDEVIAFLGEGPQRL.....KKTRDSMETENPKVLITAF 551

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR-68: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2218	78.0	557	2 JW0089	organic cation tra
2	2152	75.6	557	2 JE0346	high-affinity carn
3	754.5	26.5	576	2 T22509	hypothetical prote
4	699.5	24.6	593	2 JC4884	organic cation tra
5	681.5	24.0	556	2 S50862	organic cation tra
6	565	19.9	794	2 T27870	hypothetical prote
7	548	19.3	745	2 T16565	hypothetical prote
8	484.5	17.0	527	2 T01019	transport protein
9	475	16.7	539	2 C96758	probable protein
10	421.5	14.8	518	2 B86299	hypothetical prote
11	412.5	14.5	521	2 H86298	hypothetical prote
12	382.5	13.4	515	2 B96825	hypothetical prote
13	356	12.5	528	2 T21682	hypothetical prote
14	338	11.9	540	2 T25851	hypothetical prote
15	327	11.5	454	2 F75580	hypothetical prote
16	317	11.1	455	2 B83213	probable sugar tra
17	311	10.9	591	2 T30895	sugar transport pr
18	307	10.8	435	2 T15290	hypothetical prote
19	294.5	10.4	529	2 T23190	hypothetical prote
20	290	10.2	400	2 C69757	transporter homolo
21	286.5	10.1	478	2 T33985	hypothetical prote
22	281	9.9	422	2 G72234	hypothetical prote
23	276	9.7	461	2 D70073	metabolite transpo
24	275	9.7	520	2 T23545	hypothetical prote
25	274.5	9.6	443	2 E64725	yaaU protein - Esc
26	274.5	9.6	443	2 H85485	probable transport
27	272.5	9.6	442	2 A83122	probable MFS trans
28	271.5	9.5	524	2 T27082	hypothetical prote
29	271	9.5	423	2 T19030	hypothetical prote

ALIGNMENTS

RESULT 1

JW0089

organic cation transporter protein 2 - human

N:Alternate names: OCTN2

C:Species: Homo sapiens (man)

C:Date: 18-Jun-1998 #sequence_revision 10-Jul-1998 #text_change 28-May-1999

C:Accession: JW0089

R:Wu, X.; Prasad, P.D.; Leibach, F.H.; Ganapathy, V.

Biochem. Biophys. Res. Commun. 246, 589-595, 1998

A:Title: cDNA sequence, transport function, and genomic organization of human OCTN2, A:Reference number: JW0089; MUID:98289574

A:Accession: JW0089

A:Molecule type: mRNA

A:Residues: 1-557 <NUA>

A:Cross-references: GB:AF057164; NID:g3273740; PIDN:AAC24828.1; PID:g3273741

A:Experimental source: placenta

C:Comment: This transporter functions in the elimination of cationic drugs and other

Query Match 78.0%; Score 2218; DB 2; Length 557;

Best Local Similarity 75.9%; Pred. No. 1.6e-156;

Matches 423; Conservative 57; Mismatches 71; Indels 6; Gaps 3;

Qy 1 MRDDEVIAFLGEGPQRLIFILLASIIIPNGFNGMSVVFVLAGTPEHRCRVPDAANLSS 60

Db 1 MRDDEVIAFLGEGPQRLIFILLASIIIPNGFNGMSVVFVLAGTPEHRCRVPDAANLSS 60

Qy 61 AWRNNSVPLRLRDGREGVPHSCRYRIATANTFSAIGLEPGRDVDLGQLEQSCLDGWEFS 120

Db 61 AWRNHTVPLRLRDGREGVPHSCRYRIATANTFSAIGLEPGRDVDLGQLEQSCLDGWEFS 120

Qy 121 QDVVLSTVVTENLVCEDDNKKVPLTTSLFFVGVLLGSFVSGQSDRFGKKNVLFATMAVQ 180

Db 121 QDVVLSTVVTENLVCEDDNKKVPLTTSLFFVGVLLGSFVSGQSDRFGKKNVLFATMAVQ 180

Qy 181 TGFSLQIFSIEMFTVIVGVGQISNYVVAFTILGTEILGKSVRIIFSTLGVCTFFA 240

Db 181 TGFSLQIFSIEMFTVIVGVGQISNYVVAFTILGTEILGKSVRIIFSTLGVCTFFA 240

Qy 241 VGYMLLPFAFIFDRWMLLALTVPGLVCPVLMWFIPESPRWLIISQRFRFAEDIQKA 300

Db 241 VGYMLLPFAFIFDRWMLLALTVPGLVCPVLMWFIPESPRWLIISQRFRFAEDIQKA 300

Qy 301 AKMNTAVPAVIFD--SVEELNPLKQKAFILDLFTRNIAITMISLLMLTSGVYFA 358

Db 301 AKANGIVVPTIFDPSLEQLDKSKQSHNILLDLRTNIRMTVIMSLMWTISGVYFG 360

Qy 359 LSLDAPNLRGDAYLNCFLSALIEIPAYITAWLLRLPRYIIIAVFLWGGGVLLFIQLV 418

Db 361 LSLDTPNLHGDI FVNCFLSAMVEVPAYVLAWLLQLVLPRIYSKATALFGGSVLLFMQLV 420

Qy 419 PVDYFYSIGLVMLGKFGITSFSAFMSLVYFTAEYLPTLVLRNMAVGVTSTASRVGSIAPYF 478

probable sugar tra
hypothetical prote
hypothetical prote
hypothetical prote
hexose transport p
synaptic vesicle p
hypothetical prote
glucose transport
glucose transport p
glucose transport
glucose transport
hypothetical prote
transmembrane tran
glucose transport
glucose transporte

Db	421	PPDYLYATLVNMGKGVTAASWVYVTAELIYPTVVRNMGVGSSTASRLGSLISPYF	480
Qy	479	VYLGAYNRMLPYIVMGSLTILVIGTILFFPEGLMGLTPLEQMKQKWFNRSGK	--KTR 535
Db	481	VYLGAYDRFLPYILMGSLTILTAILFLPESFGPLDPTIDQLRVKGMKHKRTISHTP	540
Qy	536	DSMTEENPKVL - ITAF	551
Db	541	MLKDGQERPTILKSTAF	557

RESULT 2
JE0346
high-affinity carnitine transporter, CTL - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
C:Accession: JE0346
R:Sekine, T.; Kusunoha, H.; Utsunomiya-Tate, N.; Tsuda, M.; Sugiyama, Y.; Kanai, Y.; Endo
Biochem. Biophys. Res. Commun. 251, 586-591, 1998
A:Title: Molecular cloning and characterization of high-affinity carnitine transporter #
A:Reference number: JE0346; MUID:99011422
A:Accession: JE0346
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-557 <EX>
A:Cross-references: DDBJ:AB017260; NID:G3869208; PIDN:BAA34399.1; PID:G3869209

```
Query Match      75.6%; Score 2152; DB 2; Length 557;
Best Local Similarity 73.1%; Pred. No. 1.2e-151;
Matches 407; Conservative 68; Mismatches 76; Indels 6; Gaps 3;
```

[illegible][illegible][illegible]

Qy	479	VYLGAYNRMPLPYVMGSLTVLIGIFTLPFPESLGMTLFTLEQMKQVKWFRSGK	--KTR	533
Db	481	VYLGAVDREPLPYLMGSLTITLITLFLFPESFGAPLPDITDMLRVKGIQKQIQSOTR	540	
Qy	536	DSWETEENPKVL-ITAF	551	
		:		
Db	541	TQKDGESPTVLKSTAF	557	

RESULT 3
T22509
hypothetical protein F52F12.1 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T22509
R:Matthews, L.
submitted to the EMBL Data Library, December 1996
A:Reference number: Z19573
A:Accession: T22509
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-576 <MIL>
A:Cross-references: EMBL:Z83228; PIDN:CA805732.1; GSPDB:GN000019; CESP:F52F12.1
A:Experimental source: clone F52F12
C:Genetics:
A:Gene: CESP:F52F12.1
A:Map position: 1
A:Introns: 16/1; 23/3; 50/1; 80/3; 108/2; 134/3; 221/1; 257/2; 319/3; 358/2; 490/3; 5

```
Query Match      26.5% ; Score 754.5; DB 2; Length 576;
Best Local Similarity 30.2%; Pred. No. 2.8e+48;
Matches 169; Conservative 119; Mismatches 236; Indels 35; Gaps 9;
```

Qy 3 DYDEVIAFLGEMGPQRLLFFLLS-ASTIPNGFMGVVFLACTPBRHRVRVPDAANLSSA 61
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 16 DFDVLVEQMGVTGYIVFVFICLTSLPSAFSAFNIPVVGNNPHTCHIPEGKEYLRP 75
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

Qy 62 WRNNSVPLRLDRGREVPHSCSYRIATTANFSALGLEPGRDVDLGLOEQESIDQWFSQ 121
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

Db 76 LTNDTOIL-----SKCYNETQLNVRAFTSAP-VDTYSDRISLVLPQCNGWDYDN 124
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

[illegible][illegible]

Qy	354	VGVFALSDAPNLHGDAYLNCPLSALSIEIPAYITAWILLRLTPRRYIIAAVLFWGGVLL	413
Db	363	IIYNGLTNAVSNLPVDDYWSFTINGAVELPGYFVFWPLLQACAGRRWTLAATMIVCIGCV	422
Qy	414	FIQLVPVDYFFISIGLVMLGKGTISAFSMLYVFTAEIYPTFLYRNNAVGVTSFASVSI	473

[illegible]

RESULT 4
JC4884
organic cation transporter protein 2 - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 10-Sep-1996 #sequence_revision 18-Oct-1996 #text_change 05-Nov-1999
C:Accession: Jc4884

C; Accession: JC4884

C:Accession: T27870

R:White, S.

submitted to the EMBL Data Library, November 1995

A:Reference number: Z20432

A:Accession: T27870

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-794 <WIL>

A:Cross-references: EMBL:Z66567; PIDN:CAA91492.2; GSPDB:GN00028; CESP:ZK455.8

A:Experimental source: clone ZK455

C:Genetics:

A:Gene: CESP:ZK455.8

A:Map position: X

A:Introns: 52/3; 119/2; 158/1; 207/3; 335/2; 428/3; 610/2; 636/3; 733/3

Query Match 19.9%; Score 565; DB 2; Length 794;

Best Local Similarity 26.4%; Pred. No. 4.2e-34;

Matches 158; Conservative 121; Mismatches 232; Indels 88; Gaps 20;

QY 3 DYDE----VIAFIGWGPFORLIFPLLSASIPNGFMGSMVVFVLAGTPEHRCRVP----- 53

Db 155 DEEDDLGILQIGCSYQWIIIVLIISVQVPHAMFNLSVVMYQPDHCKIPFFNEE 214

QY 54 ----DAANLSSAW---RNSV--PURLRDGREVPH--SCSRYRLATIANFSALGLEPGRD 102

Db 215 SPSAELGYNTYDQVLNSTIAFPRTFNKQNELHDDCHYFE----RDYVHIKLSWQAQ 270

QY 103 V-DLQLEQESCLDGEESQDYLSTVVTENNLCEDNKKVPLTSLFFVGLGSPVSG 161

Db 271 VKDMATGKINCKEWEYDTSVMRTIIVTEWNVCDNNRAHVMSTSLGYLVGCFVGG 330

QY 162 QLSDRFGRKNVLFATMAYOTGFSFLQIFSISWEMFTVL-FVIVGQTSNTVVAIFLGE 220

Db 331 FISDRYGRKTAITGILTMFLGFLTYSKEFEIFLVVRELLAATNEAAD-LAAVVLCE 389

QY 221 ILGKSVRIESTIGVCTFFAVGYMLLPFAFYFIRDMRLALLVPGVLCVPLWFFIPES 280

Db 390 VTGTYSIVGSL--IQAPWACQAFALAIAYLTSTWMIHLICVLLLSIMLLYFLPES 448

QY 281 PRLISORFREAEIDIOAKAMNNTAVPA--VIFDSVEELNPLK--QOKAFILDLFTRN 337

Db 449 PRLNLNKTQAKIIEACHYNNKSLPSDLGLVRHAEEKKKWKHNEKPSYFHLPSSE 508

QY 338 IAIMTMSLLWN-----LTSVGY----- 356

Db 509 LRFNRNVLFIVWVIOKNIFQNSRFQYSDNSGSLRNGCYCIVRSVITRKVGRTLSEKFR 568

QY 357 -----FALSLOAFNLHGDAYLNCFLSALIEIPAYITAWLLRLTPRYIIAAVLFWGGV 411

Db 569 NOKLCFRVFD-----GNFFLNNMAGAIELPTLVFCVFLLR-MGRKRSQMLVLF-GSGL 621

QY 412 LFLIQLVPV--DYFLSLGLVGLKFGITSAFSMLYFETAEYPTLVLRNMAVGVTSTASR 469

Db 622 FLTTSVWVYRKQSTLALIFMLLSKACIQGSFNILYIFTSELNPTVVRNSAVGISSWVR 681

QY 470 VGSIIAPFVYLGAYNR-MLPIYVNGSITVLIGITFFLFFPSLGMTLPET-LEOMQVKV 526

Db 682 MGAGASGYAIIISLDVTMPLVPTIFACFSLLAGCLVLLLPETQGLPLPDTILDSVQMVK 740

RESULT 7

T16565

hypothetical protein k05f1.6 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999

C:Accession: T16565

R:Wohlmann, P.

submitted to the EMBL Data Library, June 1995

A:Description: The sequence of C. elegans cosmid K05F1.

A:Reference number: Z18537

C:Genetics:

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-745 <WOH>

A:Cross-references: EMBL:U29377; NID:g868173; PID:g868176; PIDN:AAA68713.1; CESP:K05F

A:Experimental source: strain Bristol N2

C:Genetics:

A:Gene: CESP:K05F1.6

A:Introns: 70/2; 160/3; 205/3; 249/3; 293/2; 350/1; 386/3; 462/3; 546/2; 669/3; 680/3

Query Match 19.3%; Score 548; DB 2; Length 745;

Best Local Similarity 27.8%; Pred. No. 7e-33;

Matches 166; Conservative 112; Mismatches 233; Indels 86; Gaps 21;

QY 3 DYDEVIAFLGEPWPFORLIFPLLSASIPNGFMGSMVVFVLAGTPEHRCRVP----- 53

Db 112 DFGILNIIGGCRWQIWIIVLIALQOIPHAMFNLVVYVMDYDFEQCMVPGENDTNDIT 171

QY 54 -----DAANLSSAWRN-NSVPLRLRDGREVPHSC-----SRYRLATIANFS 93

Db 172 ISSHGPVLMGVDDIKNISFVFPNANS-----DCAQYRDSCYFYERSEERYQLRRMPLE 225

QY 94 ALGLEPGRVDLQLEQESCLDGEESQDYLSTVVTENNLCEDNKKVPLTSLTSLFF-VG 152

Db 226 TAYSEAWKDV----APKKKC-QAYHFEKDVMTETIVTDENLVC-DSWFAKGAHMFYSIG 279

QY 153 VLLGSFVSGQLSDRFGRKNVLFATMAVOTGFSFLQIFSISWEMFTVL-FVIVGMGQISNTV 212

Db 280 YLLGCVLGIASDKIGRKPITIGFGILSSMLGVLPFNDYYPMFLIRLLLSAICNEAADL 339

QY 213 VAFILGTEILGKSVRIESTIGVCTFFAVGYMLLPFAFYFIRDMRLALLVPGVLCVP 272

Db 340 AAYTLCMEITGTYRAMVGSML-QAPWALGALLAIYLTSTWMIHLICVLLLSIMLLYFLPES 398

QY 273 LWFIFESPRLISORFREAEIDIOAKA-----MNNTA-----VPA---VIFDSVEE 318

Db 399 FICISPEPRLIMVQNRVSEAEVIRKACREPPFPNMCTTSCGNLPSDLELVSHREK 458

QY 319 LNLKOOKAFILDFTFRNIAIMTMSLLMLTSLVGYFALS-----DAPN---LHGDA 370

Db 459 LNKNGKGIGFLDFTMKELRYRTISVCIVFMATALVYVGLVMAISDQSAQRTLTFTGYF 518

QY 371 YLNCFLSALIEIPA-YITAWLLRLTPRYIIAAVLFWGGVLLFQLVLPVD-YVFLSIG 428

Db 519 HLANGAGAEITPLFACVMM--QLGRKKALMLTITSGLFIIVAMLSVSGHTMALA 576

QY 429 LVMLKFGITSAFSLYFETAEYPTLVLRNMAVGVTSTASRVGSIITAPYFVILGAYNRML 488

Db 577 FMYFGKIAGVQAFNLIYIFTSELYPTVVRNTAVGTSMVAREGSLSSYIALLS--NISL 634

QY 489 PYIVMGSLTVLIGITPLF-----FPESLGMTLPETLEQ-----MOKVWFRSGKKT 534

Db 635 PIVPM----IIFAVFSLAFAGMLVFLVLPETSEKPLPETLDDAINFLEPTQKFTIEST 687

RESULT 8

T01019

transport protein homolog YUP8H12R.2 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 05-Feb-1999 #sequence_revision 05-Feb-1999 #text_change 22-Oct-1999

C:Accession: T01019

R:Theologis, A.; Vysotskaya, V.S.; Osborne, B.I.; Schwartz, J.R.; Federspiel, N.A.; K

Oefner, P.; Davis, R.W.

submitted to the EMBL Data Library, May 1998

A:Description: Arabidopsis thaliana chromosome 1 YAC YUP8H12R sequence.

A:Reference number: Z14227

A:Accession: T01019

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-527 <THE>

A:Cross-references: EMBL:AC002986; NID:g2494106; PID:g3152576; GSPDB:GN000059; ATSP:YU

A:Experimental source: cultivar Columbia

C:Genetics:

A:Gene: ATSP:YUP8H12R.2


```

Db      394  ITICVFSTFYTFEYEDYSLIIIRVATLLASATASQVFTIKSVMLMEFYPTVTRNSAVS 453
Qy      463  VTSPASRVGSIAPYFVVLGAYNMLPYVMGSLTVLIGI-FTLFFPESLGMTLPETLEQ 521
Db      454  FKSSASRIGTILGQLFILCPY-KSLPYAILTGCLDAIAFQLRLPETKGKLPETMPE 512
Qy      522  MOK 524
Db      513  RHK 515

RESULT  14
T25851
hypothetical protein T01B11.5 - Caenorhabditis elegans.
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T25851
R:Geisel, C.; Stellyes, L.
submitted to the EMBL Data Library, December 1996
A:Description: The sequence of C. elegans cosmid T01B11.
A:Reference number: Z20099
A:Accession: T25851
A>Status: preliminary; translated from GE/EMBL/PDBJ
A:Molecule type: DNA
A:Residues: 1-540 <GE>
A:Cross-references: EMBL:U80931; PIDN:AAB38002.1; GSPDB:GN00022; CESP:T01B11.5
A:Experimental source: strain Bristol N2; clone T01B11
C:Genetics:
A:Gene: CESP:T01B11.5
A:Map position: 4
A:Introns: 85/1; 128/3; 209/2; 269/2; 352/2; 381/3; 489/3

Query Match      11.9%; Score 338; DB 2; Length 540;
Best Local Similarity 25.1%; Pred. No. 1.6e-17;
Matches 107; Conservative 72; Mismatches 202; Indels 46; Gaps 8

Qy      127  TVVTEWNLVCEDNWKVPLTSLFFVGVLLGSFVSGQLSDRFGKKNVLFATMAVQTGFSEL 186
Db      124  SVQVEFNICYIDAKKVKNTITVQTCGLVGAATFGVSDTFGRRKALLISTLGNGLFNII 183
Db      127  QVTEVTEWNLVCEDNWKVPLTSLFFVGVLLGSFVSGQLSDRFGKKNVLFATMAVQTGFSEL 186

```

D	b	184	TAYSPDLFYFMVWRTLAGVFTGGVTIVQVMFVVENIPRKDRMWIQN---SITWSPNLLIF' 240
Q	y	247	PLPAYFIRDWRMLLALITVPGLVCPLWNWFIPESRWLIISORFREAEADIIQAAKMNN'T 306
D	b	241	PFWAQCQRWTMSVIVAASSIAFTIACLILESPRWLIQKRIEARKSLIRIKTDK- 299
Q	y	307	AVPAVAFDSVEE-----LNPLKOOKAFILDFRFRNTAIIMTISLLIWL 351
D	b	300	----LYDETFFQLDEVLHVEAEKHARSSKKAKKYTFIHLCFTWKMIAGSLFITGITC 354
Q	y	352	TSVGYPALSIDAPNLHGDAYINC-----FLSALIETPAYITAWULLLTLPRIYIAA 403
D	b	355	TTFTIVTSLMYNMKEKLSGSLYNLAIMGASRWIINIAVSADIYKLWPFGRKWIQIAMIC- 413
Q	y	404	VLFWGCGVLLFIQLPVVDYFTLSIG-LVMJLGKGIGITSAFSML----YVFTAELYPTLVRN 458
D	b	414	-----TLFALLMIAVLYFYNGQLIAICTVAASCMCSOLFATAKMMYNELYPVTPARN 466
Q	y	459	MANGVTSTASRGSIAPYFVYLGAYNRMELPYIVMGSLTVL-IGIFTLPFPSELGMTLPE 517
D	b	467	LAVSAVSTMGRIGSMFSQLFYLSDYAEWIPYAVILFSQCYDFFVLFSLELPETKGVILEN 526
Q	y	518	TLEQMOK 524
D	b	527	HLPCHK 533
RESULT 15			
F	75580	probable sugar transporter - Deinococcus radiodurans (strain R1)	

C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
C:Accession: F75580
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
M.; Shen, M.; Vamathevan, J.O.; Lam, P.; McDonald, L.; Utterback, T.; Zaleski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896
A:Accession: F75580
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-454 <WHI>
A:Cross-references: GB:AE001863; GB:AE001825; NID:96460670; PIDN:AAF12486.1; PID:9646078
A:Experimental source: strain R1
C:Genetics:
A:Gene: DRA0271
A:Map position: 2
C:Superfamily: yaaU protein

Query Match 11.5%; Score 327; DB 2; Length 454;
Best Local Similarity 26.0%; Pred. No. 8.8e-17;
Matches 107; Conservative 81; Mismatches 166; Indels 58; Gaps 13;

QY 143 PLTTSIF---FVGLGSEVSQSDRGKKNLFTMAVQTGFSFLQIF--SISWEM-- 195
DB 67 PAATMLLTATFAGMLFGAWFGYLADRVGRKRSVELTTVALGVVFGLAGALAPLTLLVA 126
QY 196 -FTVLFVIVGMGQISNYVAFILGTEILGKSVRIIFSTGLGCTFFAVGYMLPLFYFI- 253
DB 127 RELTGAIGGTLPIVDYSMMAEFVPTAWRGFLVYLES-----FWAVGTVVVAALAWWS 180
QY 254 -----RDWRMLLLALTVPGLCVPLWFFIPESPRWLISORRFEAEEDIIQKAKMNNTA 307
DB 181 TAPAPAGWRMLLGLAALPGVLGVGIARTIGIPDSRLLARGEQAQAALQKVAQANGGT 240
QY 308 VPVAFIDSVEELNPLKQOKAFILDIFRNIAMTMSLLIWMILTSVGFALSIDAPNL- 366
DB 241 LPAAPL-----AHPQPPRVSPAQLFR-GVLARTPLLMVTWGLSLGYGIFSWLP SFL 294
QY 367 -----HGDAYLNQPLSALIEIPAYITAWLLRLTPRYIIAAVLEWGG-GVLLFTQLVP 419
DB 295 RAQGLDLGAVYRSTLLAALQVPGYLLAAYLVEKIGRRVTLVGFLTLGAVGAYLFLLAHD 354
QY 420 VDYVFLSIGVLMGKFGITSAFSLYVFTAELYPTLVNMAVGYTSTASRVGSIIAPYFV 479
DB 355 ANTLLTSALL---SFALLGAWGSLYATPELFTPLRTTGMGLVSGVARLASVWSP--- 408
QY 480 YLGAYNRMLPYIVMGSLIVLIGITFLFPF-----ESLGMTLPETLE 520
DB 409 SIGA-----MLLTGNLTALTVAFCVFAALAAAWGIGVETRGOALAEAE 454

Search completed: August 16, 2001, 13:59:27
Job time: 154 sec


```
QY 121 QDYVLTSTVTENLVCEENKWKVPLTTSFFVGVLLGSFVSGQLSDRFRGKKNVLFATMAVQ 180
|||||
Db 121 QDYVLTSTVTENLVCEENKWKVPLTTSFFVGVLLGSFVSGQLSDRFRGKKNVLFATMAVQ 180
|||||
QY 181 TGFSEFLQIFSISWEMFTVLFVIVGMGQISNYVAVFILGTEILGKSVRIIFSTLGVCTFFA 240
|||||
Db 181 TGFSEFLQIFSISWEMFTVLFVIVGMGQISNYVAVFILGTEILGKSVRIIFSTLGVCTFFA 240
|||||
QY 241 VGYMLLPFAFYFIRDRWMLLLALTVPGLVCPPLWFPISPESRWLISQRRFREAEIIQKA 300
|||||
Db 241 VGYMLLPFAFYFIRDRWMLLLALTVPGLVCPPLWFPISPESRWLISQRRFREAEIIQKA 300
|||||
QY 301 AKNNNTAVPAVIFDSVEELNPLKQKAFILDLFRTNRNIAIMTMSLLWMLTSVGYFALS 360
|||||
Db 301 AKNNNTAVPAVIFDSVEELNPLKQKAFILDLFRTNRNIAIMTMSLLWMLTSVGYFALS 360
|||||
QY 361 LDAPNLHGDAYLNCFLSALIEIPAYITAWLLRLTPRRYIIAAVLFWGGGVLLFIQLVPV 420
|||||
Db 361 LDAPNLHGDAYLNCFLSALIEIPAYITAWLLRLTPRRYIIAAVLFWGGGVLLFIQLVPV 420
|||||
QY 421 DYYFSLGVLMLGKFGITSAFSLMYVFTAEVPTLVNRNMAVGVSTASRVGSIAPFYVY 480
|||||
Db 421 DYYFSLGVLMLGKFGITSAFSLMYVFTAEVPTLVNRNMAVGVSTASRVGSIAPFYVY 480
|||||
QY 481 LGAYNRMLPYIVNGSLTVLIGITLFFPESLGMTLPETLEQMKVKWFRSGKTRDSMET 540
|||||
Db 481 LGAYNRMLPYIVNGSLTVLIGITLFFPESLGMTLPETLEQMKVKWFRSGKTRDSMET 540
|||||
QY 541 EENPKVLITAF 551
|||||
Db 541 EENPKVLITAF 551
|||||

RESULT 2
Q9H015
ID Q9H015 PRELIMINARY; PRT; 551 AA.
AC Q9H015
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE INTEGRAL MEMBRANE TRANSPORT PROTEIN.
GN UT2H.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Spritzberger F., Gruendemann D., Schoemig E.;
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; Y09881; CAA71007.1; -.
SQ SEQUENCE 551 AA; 62155 MW; C827A99AA78C9443 CRC64;

Query Match 99.6%; Score 2833; DB 4; Length 551;
Best Local Similarity 99.6%; Pred. No. 6.8e-165;
Matches 549; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MRDYDEVIAFLGEGWGPQRLIFLLSASIIIPNGFNMGSVVFLAGTPEHRCRVDAANLSS 60
|||||
Db 1 MRDYDEVIAFLGEGWGPQRLIFLLSASIIIPNGFNMGSVVFLAGTPEHRCRVDAANLSS 60
|||||
QY 61 AWRNNSVPLRLRDGRVPHSCSRYRLATIANFSALGLEPGRDVLGQLEQSCLDGWEEFS 120
|||||
Db 61 AWRNNSVPLRLRDGRVPHSCSRYRLATIANFSALGLEPGRDVLGQLEQSCLDGWEEFS 120
|||||
QY 121 QDYVLTSTVTENLVCEENKWKVPLTTSFFVGVLLGSFVSGQLSDRFRGKKNVLFATMAVQ 180
|||||
Db 121 QDYVLTSTVTENLVCEENKWKVPLTTSFFVGVLLGSFVSGQLSDRFRGKKNVLFATMAVQ 180
|||||
QY 181 TGFSEFLQIFSISWEMFTVLFVIVGMGQISNYVAVFILGTEILGKSVRIIFSTLGVCTFFA 240
|||||
```

```
Db 181 TGFSEFLQIFSISWEMFTVLFVIVGMGQISNYVAVFILGTEILGKSVRIIFSTLGVCTFFA 240
QY 241 VGYMLLPFAFYFIRDRWMLLLALTVPGLVCPPLWFPISPESRWLISQRRFREAEIIQKA 300
|||||
Db 241 VGYMLLPFAFYFIRDRWMLLLALTVPGLVCPPLWFPISPESRWLISQRRFREAEIIQKA 300
|||||
QY 301 AKNNNTAVPAVIFDSVEELNPLKQKAFILDLFRTNRNIAIMTMSLLWMLTSVGYFALS 360
|||||
Db 301 AKNNNTAVPAVIFDSVEELNPLKQKAFILDLFRTNRNIAIMTMSLLWMLTSVGYFALS 360
|||||
QY 361 LDAPNLHGDAYLNCFLSALIEIPAYITAWLLRLTPRRYIIAAVLFWGGGVLLFIQLVPV 420
|||||
Db 361 LDAPNLHGDAYLNCFLSALIEIPAYITAWLLRLTPRRYIIAAVLFWGGGVLLFIQLVPV 420
|||||
QY 421 DYYFSLGVLMLGKFGITSAFSLMYVFTAEVPTLVNRNMAVGVSTASRVGSIAPFYVY 480
|||||
Db 421 DYYFSLGVLMLGKFGITSAFSLMYVFTAEVPTLVNRNMAVGVSTASRVGSIAPFYVY 480
|||||
QY 481 LGAYNRMLPYIVNGSLTVLIGITLFFPESLGMTLPETLEQMKVKWFRSGKTRDSMET 540
|||||
Db 481 LGAYNRMLPYIVNGSLTVLIGITLFFPESLGMTLPETLEQMKVKWFRSGKTRDSMET 540
|||||
QY 541 EENPKVLITAF 551
|||||
Db 541 EENPKVLITAF 551
|||||

RESULT 3
Q9R141
ID Q9R141 PRELIMINARY; PRT; 553 AA.
AC Q9R141
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE ORGANIC CATION TRANSPORTER OCTN1.
GN OCTN1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=101116;
RN [1]
RP SEQUENCE FROM N.A.
RA Wu X., Wang H., Leibach F.H., Ganapathy V.;
RT "Functional characteristics and tissue distribution pattern of OCTN1,
RT an organic cation transporter";
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AFL69831; AAD46922.1; -.
DR InterPro; IPR001066; -.
DR Pfam; PF00083; sugar_tr; 1.
DR PROSITE; PS00216; SUGAR_TRANSPORT_1; UNKNOWN_1.
SQ SEQUENCE 553 AA; 62362 MW; E26C8155768A14AD CRC64;

Query Match 87.3%; Score 2484; DB 11; Length 553;
Best Local Similarity 85.2%; Pred. No. 1.2e-143;
Matches 471; Conservative 36; Mismatches 44; Indels 2; Gaps 1;

QY 1 MRDYDEVIAFLGEGWGPQRLIFLLSASIIIPNGFNMGSVVFLAGTPEHRCRVDAANLSS 60
|||||
Db 1 MRDYDEVIAFLGEGWGPQRLIFLLSASIIIPNGFNMGSVVFLAGTPEHRCRVDAANLSS 60
|||||
QY 61 AWRNNSVPLRLRDGRVPHSCSRYRLATIANFSALGLEPGRDVLGQLEQSCLDGWEEFS 120
|||||
Db 61 AWRNNSVPLRLRDGRVPHSCSRYRLATIANFSALGLEPGRDVLGQLEQSCLDGWEEFS 120
|||||
QY 121 QDYVLTSTVTENLVCEENKWKVPLTTSFFVGVLLGSFVSGQLSDRFRGKKNVLFATMAVQ 180
|||||
Db 121 QDYVLTSTVTENLVCEENKWKVPLTTSFFVGVLLGSFVSGQLSDRFRGKKNVLFATMAVQ 180
|||||
QY 181 TGFSEFLQIFSISWEMFTVLFVIVGMGQISNYVAVFILGTEILGKSVRIIFSTLGVCTFFA 240
|||||
Db 181 TGFSEFLQIFSISWEMFTVLFVIVGMGQISNYVAVFILGTEILGKSVRIIFSTLGVCTFFA 240
|||||
```



```

GN OCr-1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99227113; PubMed=10209228;
RA Wu X., Fei Y.J., Huang W., Chancy C., Leibach F.H., Ganapathy V.;
RT "Identity of the F52F12.1 gene product in Caenorhabditis elegans as an
RT organic cation transporter.";
RL Biochim. Biophys. Acta 1418:239-244(1999).
CC -|- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -|- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.
DR EMBL; AF110415; AAF21932.1; -.
DR InterPro; IPR001066; -.
DR Pfam; PF00083; sugar_tr; 1.
KW Transmembrane.
SQ SEQUENCE 568 AA; 63505 MW; COA3E73851F44056 CRC64;

Query Match 26.5%; Score 754.5; DB 5; Length 568;
Best Local Similarity 30.2%; Pred. No. 1.9e-38;
Matches 169; Conservative 119; Mismatches 236; Indels 35; Gaps 9;

QY 3 DYDEVIATFLGEMGPQRILFFLLS-ASIIIPNGFMGSMVFLAGTPEHRCRVPDANLSSA 61
DB 8 DDFVLEQNGNYGTQIVFFFIICLTSLPSAFSAFNPFFVGNPPHTCHTPEGEKYLRP 67
QY 62 WRNNSVPLRLDRGREVPHSCSRYLATIANFSALGLEPGRVDLQOLBOESCLDGFESQ 121
DB 68 LTNDTQIL-----SKQYNETQINVFRAFTSAP-VDTYSDRISLVPCQNGWDYDN 116
QY 122 DYILSTVVTWNLVCEDNKWKVPLTTSFLFVGLGSGFSYSGQLSDRFRGNKLVLFATMAVQT 181
DB 117 STYLSLVTEFNLCVDAQWIEISTTSFYVSGFIGNCLFGYVADKFGRRRSFFVILTTLI 176
QY 182 GFSFLQIFSIWEMTVLFIIVGMCQISNVVAFILGTEILGKSVRIIFSLGVCVTFEAV 241
DB 177 VCGTASSFAKDIESFIILRFTFTGLAFFALQIPFIICMEFMGNSGR-IFSGMLTSFFGA 235
QY 242 GYMLLPFAFYFIRDRMILLALTVPGLVCLPLWFIPESPRWLISORRFRFREAEDIIQAA 301
DB 236 AMALGVVAMFIRRRQLTFFCNAPFAFYIIYFFLPESPRWSVSGKWADAKKQLKIA 295
QY 302 KMN---NTAVPAVIFDSVEELNPLKQKAF-----ILDFTFRNTAINTIMSLMLMITS 353
DB 296 KMGKSNVDVDELV-DSMKNHQNAAEKETKRSNHNVDLFTKPNLRRKTLIVTYIWMNA 354
QY 354 VGYFALSDAPNLHGDAYLNCFLSALIEIPAYITAWLLRLTPRYIIAANLVFWGGVLL 413
DB 355 IYNGLTNLVSNLPVDVYWSFIINGAVELPGYFVVMVPLLCACGRRTWLAATMIVCGIGCV 414
QY 414 FIQLVPDYYFISGLVLMGKFGITSAFSLMYVFTAEIYPTLVRNMAVGVSTASRVGSI 473
DB 415 SAMFMPDGYPLWLVASAFSGKGVSGFAVIYIFAGELYPTVVRVRAIGMSSWAGSGLL 474
QY 474 IAPYFVYLGAYNRMLPYIVMGSLSLVIGITFLFFPESIGMTLPETLEOMQKVKFRSGKK 533
DB 475 LAPHIVNLGKIVKILPLLINGLMALSAGILTFELPETLGAPLPMTIEDAENP-----GKK 529
QY 534 -----TRDSMETEENP 544
DB 530 PEPDSGMFTQAARKRESQP 548

RESULT 11
ID O02270 PRELIMINARY; PRT; 576 AA.
AC O02270;
DT 01-JUL-1997 (TREMBlrel. 04, Created)
DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)

```

```

DE F52F12.1 PROTEIN.
GN F52F12.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
RA Smalton N., Smith A., Sonnhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
CC -|- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -|- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.
DR EMBL; Z83228; CAB05732.1; -.
DR InterPro; IPR001066; -.
DR Pfam; PF00083; sugar_tr; 1.
KW Transmembrane.
SQ SEQUENCE 576 AA; 64493 MW; F494EE94A7EBC0B1 CRC64;

Query Match 26.5%; Score 754.5; DB 5; Length 576;
Best Local Similarity 30.2%; Pred. No. 1.9e-38;
Matches 169; Conservative 119; Mismatches 236; Indels 35; Gaps 9;

QY 3 DYDEVIATFLGEMGPQRILFFLLS-ASIIIPNGFMGSMVFLAGTPEHRCRVPDANLSSA 61
DB 16 DDFVLEQNGNYGTQIVFFFIICLTSLPSAFSAFNPFFVGNPPHTCHTPEGEKYLRP 75
QY 62 WRNNSVPLRLDRGREVPHSCSRYLATIANFSALGLEPGRVDLQOLBOESCLDGFESQ 121
DB 76 LTNDTQIL-----SKQYNETQINVFRAFTSAP-VDTYSDRISLVPCQNGWDYDN 124
QY 122 DYILSTVVTWNLVCEDNKWKVPLTTSFLFVGLGSGFSYSGQLSDRFRGNKLVLFATMAVQT 181
DB 125 STYLSLVTEFNLCVDAQWIEISTTSFYVSGFIGNCLFGYVADKFGRRRSFFVILTTLI 184
QY 182 GFSFLQIFSIWEMTVLFIIVGMCQISNVVAFILGTEILGKSVRIIFSLGVCVTFEAV 241
DB 185 VCGTASSFAKDIESFIILRFTFTGLAFFALQIPFIICMEFMGNSGR-IFSGMLTSFFGA 243
QY 242 GYMLLPFAFYFIRDRMILLALTVPGLVCLPLWFIPESPRWLISORRFRFREAEDIIQAA 301
DB 244 AMALGVVAMFIRRRQLTFFCNAPFAFYIIYFFLPESPRWSVSGKWADAKKQLKIA 303
QY 302 KMN---NTAVPAVIFDSVEELNPLKQKAF-----ILDFTFRNTAINTIMSLMLMITS 353
DB 304 KMGKSNVDVDELV-DSMKNHQNAAEKETKRSNHNVDLFTKPNLRRKTLIVTYIWMNA 362
QY 354 VGYFALSDAPNLHGDAYLNCFLSALIEIPAYITAWLLRLTPRYIIAANLVFWGGVLL 413
DB 363 IYNGLTNLVSNLPVDVYWSFIINGAVELPGYFVVMVPLLCACGRRTWLAATMIVCGIGCV 422
QY 414 FIQLVPDYYFISGLVLMGKFGITSAFSLMYVFTAEIYPTLVRNMAVGVSTASRVGSI 473
DB 423 SAMFMPDGYPLWLVASAFSGKGVSGFAVIYIFAGELYPTVVRVRAIGMSSWAGSGLL 482
QY 474 IAPYFVYLGAYNRMLPYIVMGSLSLVIGITFLFFPESIGMTLPETLEOMQKVKFRSGKK 533

```


Qy	1	MRDYDEVIAFLGEGWGPQRLIFF---LLSASIIPNCFNGMSVVFELAGTPEHCRCPVDDAAN	57
Db	1	MTAVDDILEHTGEGFNFQKOTFELLALLSAAFTTP---IYVGIVLGFIPDHCRSPGVAAE	57
Qy	58	LSS---AMR-----NNSVPLRLDRGVPVHSCSRYLATIANFSALG--LEP--GROVDLQO	107
Db	58	LSQRCGSLAAELNYTPVPGPAGAPFQRCRRYEDVW--NQSTLCGVPLAGLAANSSH	115
Qy	108	LEQESCLDGHFEFSQDYVLSVTVTEWNLICEDNNKVPILTTSLEFVGVLLGSFVSGOLSDFR	167
Db	116	LPUGPCRYGWY---DTPGSSIVTFEFLVCANSMLLDLFOSAVNVGFFIGSVGIGYIADFRL	173
Qy	168	GRKNVLPATMAVOTGFSEFLQIFSISSWEMFTVLEFVIVGMQGISYNYVAVFALTGTELGKSVR	227
Db	174	GRKLCLLLTILINAVSGVLMASPTVTWMLVFRLIQGLVSKAGWMGIYTLITFEFVGLSYR	233
Qy	228	IIFSTLGVG--TFPANGYMLLPFLFAFETDRWRLLALTVPGLCVPLMWFIPESPRLII	285
Db	234	---RTVGIFVQVAFTEFGLLVAGVAYALPHWRMLQFTVILPNECFYIWCVPESPRLII	290
Qy	286	SQRFRFAEEDIQKAAKNNTAVPAVIFDSVEELNPLQO---OKAFILDLDFLTRNIAM	341
Db	291	SQNKNAKAMSTIIKHIAKNGKSLPA-----SLOSIRPDEEVEGKLSFGLDVLTPQIRKH	346

Qy	1	MRD	DEVIA	FLG	EWG	PQR	LIF	FLS	AS	I	P	N	G	F	N	G	-	-	NSV	V	L	A	C	T	P	H	R	C	R	V	P	D	A	N	L	58	
Db	1	MST	VDD	I	L	E	H	I	G	E	H	L	F	P	K	O	T	F	L	-	-	A	L	S	G	A	F	T	P	I	Y	G	I	V	E	L	58
Qy	59	SS	-	A	W	R	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	103		
Db	59	SQR	G	W	S	O	A	E	L	N	T	V	P	C	L	G	P	S	D	E	A	S	L	S	O	C	M	R	E	V	D	N	O	S	T	L	112
Qy	104	D	L	G	I	O	E	O	E	S	C	L	D	G	S	Q	D	V	Y	L	S	T	V	T	E	N	L	V	E	D	N	K	K	V	P	L	163
Db	113	D	R	N	L	P	L	G	P	C	E	H	G	W	Y	N	T	P	G	-	-	S	S	I	V	E	F	N	L	V	A	C	H	M	L	170	
Qy	164	S	D	R	G	R	K	N	V	L	F	A	T	M	A	V	Q	T	G	S	F	L	O	I	F	S	I	S	E	M	E	T	V	L	F	V	223
Db	171	A	D	R	G	R	K	L	V	T	I	N	A	I	S	G	A	M	A	I	S	P	N	A	M	L	V	F	R	F	L	G	L	G	V	S	230
Qy	224	K	S	V	R	I	F	S	T	L	G	V	-	-	T	F	A	V	G	Y	M	L	P	L	F	A	I	R	D	R	M	L	L	A	V	P	281
Db	231	L	G	Y	-	-	-	R	M	V	G	I	O	I	A	T	V	G	L	L	I	A	G	V	A	V	I	P	N	R	M	L	O	F	A	V	287
Qy	282	R	W	L	S	O	R	F	R	E	A	E	I	O	K	A	R	N	N	T	A	P	A	V	I	F	D	S	V	E	E	L	N	P	L	K	341
Db	341	R	W	L	S	O	R	F	R	E	A	E	I	O	K	A	R	N	N	T	A	P	A	V	I	F	D	S	V	E	E	L	N	P	L	K	341

Db 288 RWLISQNKIVKAMKIIKHIAKNGKSVPSLQNLTPDEDAGKKLKPSEILDVVRTPQIRKH 347
Qy 342 TIMSLLWMLTSGVYFALSLDAPNLHGD-AYLNCFLSALIEIPAYITAMLLRLTLPYVI 400
Db 348 TLILMYNFTSSVLYXQGLIHHM-GLAGDNIYLDFFYSALVEFFPAFIIITIDRVGRYP 406
Qy 401 IAAVLFWGGVLLFPQLVPDYVYFSLGVLMLGKFGITSFMSLYVFTAELYPTLVNRMA 460
Db 407 WAVSNMVAGAACTASVFIPDDQLWKITIACLGRMGITMAYEMVCLVNAELYPTYIRNLG 466
Qy 461 VGVSTASRVGSIITAPYEVY-LGAYNRMPLPIVNGSLTVLIGIFTLFFPESLGMTLPETL 519
Db 467 VLVCSMCDIGIITPFLVYKLTIDWMEFFLVVFAVVGIVAGALVLLLPETRGKALPETI 526
Qy 520 EQMQKVKWFRSGKTR 535
Db 527 EDAENMQ--RPRKKER 540

Search completed: August 16, 2001, 14:03:35
Job time: 307 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 16, 2001, 13:58:58 ; Search time 18.15 Seconds
(without alignments)
1039.932 Million cell updates/sec

Title: US-09-521-195-1

Perfect score: 2845

Sequence: 1 MRDYDEVIAFLGNGPQRL.....KKTRDSMETEENPKVLITAF 551

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2218	78.0	557	1	OCN2_HUMAN
2	2165	76.1	557	1	OCN2_RAT
3	2164	76.1	557	1	OCN2_MOUSE
4	393	75.1	557	1	YLX5_CAEEL
5	317.5	11.2	1222	1	YMP3_CAEEL
6	307	10.8	435	1	YT13_CAEEL
7	294.5	10.4	529	1	YOL1_CAEEL
8	290	10.2	400	1	YCE1_BACSU
9	274.5	9.6	443	1	YAAU_ECOLI
10	270	9.5	461	1	CSBC_BACSU
11	262	9.2	445	1	YGCS_ECOLI
12	248	8.7	490	1	GTR1_CHICK
13	247	8.7	592	1	HXT5_YEAST
14	245	8.6	459	1	YDJK_ECOLI
15	245	8.6	566	1	HKT2_KLULA
16	244.5	8.6	567	1	HXT9_YEAST
17	243	8.5	451	1	GTR1_PIG
18	243	8.5	492	1	GTR1_RAT
19	240	8.4	495	1	GTR3_CANFA
20	238	8.4	492	1	GTR1_BOVIN
21	238	8.4	492	1	GTR1_MOUSE
22	237.5	8.3	452	1	YDJE_ECOLI
23	237	8.3	492	1	GTR1_HUMAN
24	237	8.3	546	1	HXT0_YEAST
25	237	8.3	567	1	HXTA_YEAST
26	236	8.3	451	1	YJAJ_BACSU
27	234.5	8.2	494	1	GTR3_SHEEP
28	234	8.2	491	1	XYLE_ECOLI
29	231	8.1	522	1	GTR2_RAT
30	226	7.9	457	1	YIR0_YEAST
31	225	7.9	492	1	GTR1_RABIT
32	225	7.9	763	1	RGF2_YEAST
33	223	7.8	413	1	MUCK_ACICA

34	222	7.8	523	1	GTR2_MOUSE	P14246 mus musculus
35	222	7.8	742	1	SYV2_RAT	Q02563 rattus norv
36	221.5	7.8	570	1	HXT6_YEAST	P39003 saccharomyc
37	221.5	7.8	570	1	HXT7_YEAST	P39004 saccharomyc
38	220	7.7	496	1	GTR3_HUMAN	P11169 homo sapien
39	219.5	7.7	472	1	ARAE_KLEOX	P45598 klebsiella
40	218	7.7	569	1	HXT8_YEAST	P40886 saccharomyc
41	215.5	7.6	472	1	ARAE_ECOLI	P09830 escherichia
42	215.5	7.6	551	1	HGT1_KLULA	P49374 kluyveromyc
43	215	7.6	448	1	PCAK_PSEPU	Q51955 pseudomonas
44	215	7.6	457	1	PCAK_ACICA	Q43975 acinetobact
45	215	7.6	493	1	GTR3_MOUSE	P32037 mus musculu

ALIGNMENTS

RESULT 1						
OCN2_HUMAN						
ID	OCN2_HUMAN	STANDARD;	PRT;	557 AA.		
AC	O76082;					
DT	01-OCT-2000 (Rel. 40, Created)					
DT	01-OCT-2000 (Rel. 40, Last sequence update)					
DT	01-OCT-2000 (Rel. 40, Last annotation update)					
DE	ORGANIC CATION/CARNITINE TRANSPORTER 2 (SOLUTE CARRIER FAMILY 22,					
DE	MEMBER 5) (HIGH-AFFINITY SODIUM-DEPENDENT CARNITINE COTRANSPORTER).					
GN	SLC22A5 OR OCTN2.					
OS	Homo sapiens (Human).					
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;					
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.					
OX	NCBI_taxid=9606;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RX	MEDLINE=98289574; PubMed=9618255;					
RA	Wu X., Prasad P.D., Leibach F.H., Ganapathy V.;					
RT	"cDNA sequence, transport function, and genomic organization of human					
RT	OCTN2, a new member of the organic cation transporter family.";					
RL	Biochem. Biophys. Res. Commun. 246:589-595(1998).					
RN	[2]					
RP	SEQUENCE FROM N.A.					
RC	TISSUE=Kidney;					
RX	MEDLINE=98352077; PubMed=9685390;					
RA	Tamai I., Ohashi R., Nezu J.-I., Yabuuchi H., Oku A., Shimane M.,					
RA	Sai Y., Tsuji A.;					
RT	"Molecular and functional identification of sodium ion-dependent, high					
RT	affinity human carnitine transporter OCTN2.";					
RL	J. Biol. Chem. 273:20378-20382(1998).					
RN	[3]					
RP	SEQUENCE FROM N.A.					
RX	MEDLINE=99113835; PubMed=9916797;					
RA	Nezu J., Tamai I., Oku A., Ohashi R., Yabuuchi H., Hashimoto N.,					
RA	Nikaido H., Sai Y., Koizumi A., Shoji Y., Takada G., Matsuishi T.,					
RA	Yashino M., Kato H., Ohura T., Tsujimoto G., Hayakawa J., Shimane M.,					
RA	Tsuji A.;					
RT	"Primary systemic carnitine deficiency is caused by mutations in a					
RT	gene encoding sodium ion-dependent carnitine transporter.";					
RL	Nat. Genet. 21:91-94(1999).					
RN	[4]					
RP	CHARACTERIZATION.					
RX	MEDLINE=99384224; PubMed=10454528;					
RA	Wu X., Huang W., Prasad P.D., Seth P., Rajan D.P., Leibach F.H.,					
RA	Chen J., Conway S.J., Ganapathy V.;					
RT	"Functional characteristics and tissue distribution pattern of organic					
RT	cation transporter 2 (OCTN2), an organic cation/carnitine					
RT	transporter.";					
RL	J. Pharmacol. Exp. Ther. 290:1482-1492(1999).					
RN	[5]					
RP	VARIANT CDS GLN-169.					
RX	MEDLINE=99355597; PubMed=10425211;					
RA	Burwinkel B., Kreuder J., Schweitzer S., Vorgerd M., Gempel K.,					
RA	Gerbitz K.-D., Killmann M.W.;					
RT	"Carnitine transporter OCTN2 mutations in systemic primary carnitine					
RT	deficiency: a novel Arg169Gln mutation and a recurrent Arg282ter					

mutation associated with an unconventional splicing abnormality.";
 Biochem. Biophys. Res. Commun. 261:484-487(1999).
 [6]
 RX MEDLINE-99408248; PubMed-10480371;
 RA Vaz F.M., Scholte H.R., Rutter J., Hussaarts-Odijk L.M.,
 RA Rodrigues Pereira R., Schweitzer S., de Klerk J.B.C., Waterham H.R.,
 RA Wanders R.J.A.;
 RT "Identification of two novel mutations in OCTN2 of three patients with
 RT systemic carnitine deficiency";
 RL Hum. Genet. 105:157-161(1999).
 [7]
 RP VARIANT CDSP LEU-478.
 RX MEDLINE-99172075; PubMed-10072434;
 RA Tang N.L., Ganapathy V., Wu X., Hui J., Seth P., Yuen P.M.,
 RA Wanders R.J., Fok T.F., Hjeltn N.M.;
 RT "Mutations of OCTN2, an organic cation/carnitine transporter, lead to
 RT deficient cellular carnitine uptake in primary carnitine deficiency";
 RL Hum. Mol. Genet. 8:655-660(1999).
 [8]
 RN CHARACTERIZATION OF VARIANT CDSP LEU-478, AND MUTAGENESIS.
 RX MEDLINE-20026865; PubMed-10559218;
 RA Seth P., Wu X., Huang W., Leibach F.H., Ganapathy V.;
 RT "Mutations in novel organic cation transporter (OCTN2), an organic
 RT cation/carnitine transporter, with differential effects on the
 RT organic cation transport function and the carnitine transport
 RT function";
 RL J. Biol. Chem. 274:33388-33392(1999).
 [9]
 RN VARIANTS CDSP ARG-283 AND PHE-446.
 RX MEDLINE-20081068; PubMed-10612840;
 RA Mayatepek E., Nezu J., Tamai I., Oku A., Katsura M., Shimane M.,
 RA Tsuji A.;
 RT "Two novel missense mutations of the OCTN2 gene (W283R and V446F) in a
 RT patient with primary systemic carnitine deficiency";
 RL Hum. Mutat. 15:118-118(2000).
 [10]
 RN VARIANT CDSP LYS-452.
 RX MEDLINE-20145665; PubMed-10679939;
 RA Wang Y., Kelly M.A., Cowan T.M., Longo N.;
 RT "A missense mutation in the OCTN2 gene associated with residual
 RT carnitine transport activity";
 RL Hum. Mutat. 15:238-245(2000).
 CC -1- FUNCTION: SODIUM-ION DEPENDENT, HIGH AFFINITY CARNITINE
 CC TRANSPORTER. ALSO TRANSPORTS ORGANIC CATIONS WITHOUT THE
 CC INVOLVEMENT OF SODIUM. INVOLVED IN THE ACTIVE CELLULAR UPTAKE OF
 CC CARNITINE.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -1- TISSUE SPECIFICITY: STRONGLY EXPRESSED IN KIDNEY, SKELETAL MUSCLE,
 CC HEART AND PLACENTA.
 CC -1- DISEASE: DEFECTS IN SLC22A5 ARE THE CAUSE OF SYSTEMIC PRIMARY
 CC CARNITINE DEFICIENCY (CDSP). CDSP IS AN AUTOSOMAL RECESSIVE
 CC DISORDER OF FATTY ACID OXIDATION CAUSED BY DEFECTIVE CARNITINE
 CC TRANSPORT. PRESENT EARLY IN LIFE WITH HYPOKETOTIC HYPOGLYCEMIA AND
 CC ACUTE METABOLIC DECOMPENSATION, OR LATER IN LIFE WITH SKELETAL
 CC MYOPATHY OR CARDIOMYOPATHY.
 CC -1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY. ORGANIC
 CC CATION SUBFAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; AF057164; AAC24828.1; -;
 DR EMBL; AB015050; BAA29023.1; -;
 DR EMBL; AB016625; BAA36712.1; -;
 DR MIM; 603377; -;
 DR MIM; 212140; -;
 DR InterPro: IPR001066; -;

DR InterPro: IPR001687; -;
 DR Pfam; PF00083; sugar_tr; 1.
 DR PROSITE; PS00216; SUGAR_TRANSPORT_1; 1.
 KW Transport; Transmembrane; Glycoprotein; Disease mutation.
 FT TRANSMEM 21 163
 FT TRANSMEM 143 163
 FT TRANSMEM 173 193
 FT TRANSMEM 196 216
 FT TRANSMEM 233 253
 FT TRANSMEM 258 278
 FT TRANSMEM 343 363
 FT TRANSMEM 372 392
 FT TRANSMEM 414 434
 FT TRANSMEM 437 457
 FT TRANSMEM 489 509
 FT CARBOHYD 57 57
 FT CARBOHYD 64 64
 FT CARBOHYD 91 91
 FT VARIANT 169 169
 FT VARIANT 211 211
 FT VARIANT 283 283
 FT VARIANT 446 446
 FT VARIANT 452 452
 FT VARIANT 478 478
 FT MUTAGEN 352 352
 FT SEQUENCE 557 AA; 62751 MW; 928B1F6EFFG3C48D CRC64;
 SQ
 Query Match 78.0%; Score 2218; DB 1; Length 557;
 Best Local Similarity 75.9%; Pred. No. 4.2e-141;
 Matches 423; Conservative 57; Mismatches 71; Indels 6; Gaps 3;
 Qy 1 MRDYDEVIATFGEWGPQRLLFFLLSIIIPNGFNGMSVVFVLAGTPHRCRVDPDAANLSS 60
 Db 1 MRDYDEVIATFGEWGPQRLLFFLLSIIIPNGFNGMSVVFVLAGTPHRCRVDPDAANLSS 60
 Qy 61 AWRNNSVPLRLDRDREPHSCRYRLATIANFSALEPGRDVDLQLEQESCLDGWEFS 120
 Db 61 AWRNHTVPLRLDRDREPHSCRYRLATIANFSALEPGRDVDLQLEQESCLDGWEFS 120
 Qy 121 QDVYLSVTWTEWNLVCEENKVKPLTTSILFFVGVLLGSFVSGQLSDREGRKNVLFATMAVQ 180
 Db 121 QDVYLSVTWTEWNLVCEENKVKPLTTSILFFVGVLLGSFVSGQLSDREGRKNVLFATMAVQ 180
 Qy 181 TGFSLQIFTSWEMFTVLFVIVGVGQISNVVAFILGTEILGSKSVRIIFSTLGVCTFFA 240
 Db 181 TGFSLQIFSKNFEMFVFLVGVGQISNVVAFVILGTEILGSKSVRIIFSTLGVCTFFA 240
 Qy 241 VGYMLLPFAFIRDRWMLLALTPGVLCVPLWMTFPESPRMLISORRFEAEIIOKA 300
 Db 241 FGMYVLPFAFIRDRWMLLALTPGVLCVPLWMTFPESPRMLISORRFEAEIIOKA 300
 Qy 301 AKMNTAVPAVIFD--SVEELNPLKQOKAFIDLFRTRNIAIMTISLLWMLTSGVYFA 358
 Db 301 AKANGIVVPSTIFDSELDLSSKKQSHNLDLRTNIRWMTIMSLMWTISVGYFG 360
 Qy 359 LSLDAPNLHGDAYNCLFSLALIEIPAYITAWLLRLTPRYITAAVLFWGGVLLFLQLV 418
 Db 361 LSLDTPNLHGDIYFNCFLSANVEVPAYVLAWLLQLYLPFRYSWATALLFLGSGVLLFNQLV 420
 Qy 419 PVDYFELSIGLVMLGKFGITSAFSLMYVFAEYLPVLVNRNAVGVTTASRGVSIAPYF 478
 Db 421 PPDLYLATVLVWVGKFGVTAFAFMSVYVYTAELPFTVVRNMGVSVSTASRLSILSPYF 480

QY 479 VYLGAYNMLPYVWGLSVLVLIGITLFPESLGMTLPETLEOMQKVKWFRSGK---KTR 535
 DB 481 VYLGAYDRPLPYLWGLSVLVLITLFLPESFGTLPDIDQMLRVKGMKHKRTPSHTR 540
 QY 536 DSMETEENPKVL-ITAF 551
 DB 541 MLKDGQERTILKSTAF 557

RESULT 2
 OCN2_RAT
 ID OCN2_RAT STANDARD; PRT; 557 AA.
 AC O70594; O90WLO;
 DT 01-OCT-2000 (Rel. 40, Created)
 DT 01-OCT-2000 (Rel. 40, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE ORGANIC CATION/CARNITINE TRANSPORTER 2 (SOLUTE CARRIER FAMILY 22,
 DE MEMBER 5) (HIGH-AFFINITY SODIUM-DEPENDENT CARNITINE COTRANSPORTER)
 DE (USTR2R) (CTL)
 GN SLC22A5 OR OCTN2.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Kidney;
 RX MEDLINE=98200080; PubMed=9541011;
 RA Schoenig E., Spitzberger F., Engelhardt M., Martel F., Oerding N.,
 RA Gruendemann D.;
 RT "Molecular cloning and characterization of two novel transport
 RT proteins from rat kidney";
 RL FEBS Lett. 425:79-86(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Intestine;
 RX MEDLINE=99011422; PubMed=9792817;
 RA Sekine T., Kusuha H., Utsunomiya-Tate N., Tsuda M., Sugiyama Y.,
 RA Kanai Y., Endou H.;
 RT "Molecular cloning and characterization of high-affinity carnitine
 RT transporter from rat intestine";
 RL Biochem. Biophys. Res. Commun. 251:586-591(1998).
 RN [3]
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
 RX MEDLINE=99384224; PubMed=10454528;
 RA Wu X., Huang W., Prasad P.D., Seth P., Rajan D.P., Leibach F.H.,
 RA Chen J., Conway S.J., Ganapathy V.;
 RT "Functional characteristics and tissue distribution pattern of organic
 RT cation transporter 2 (OCTN2), an organic cation/carnitine
 RT transporter";
 RL J. Pharmacol. Exp. Ther. 290:1482-1492(1999).
 CC -1- FUNCTION: SODIUM-ION DEPENDENT, HIGH AFFINITY CARNITINE
 CC TRANSPORTER. ALSO TRANSPORTS ORGANIC CATIONS WITHOUT THE
 CC INVOLVEMENT OF SODIUM. INVOLVED IN THE ACTIVE CELLULAR UPTAKE OF
 CC CARNITINE.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN THE PROXIMAL AND DISTAL TUBULES
 CC AND IN THE GLOMERULI IN THE KIDNEY, IN THE MYOCARDIUM, VALVES, AND
 CC ARTERIOLES IN THE HEART, IN THE LABYRINTHINE LAYER OF THE
 CC PLACENTA, AND IN THE CORTEX, HIPPOCAMPUS, AND CEREBELLUM IN THE
 CC BRAIN.
 CC -1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY. ORGANIC
 CC CATION SUBFAMILY.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 EMBL; AJ001933; CAA05106.1; -

DR EMBL; AB017260; BAA34399.1; -
 DR EMBL; AF110416; AAD54059.1; -
 DR InterPro; IPR001066; -
 DR Pfam; PF00083; sugar_tr; 1.
 DR PROSITE; PS00216; SUGAR_TRANSPORT_1; 1.
 KW Transport; Transmembrane; Glycoprotein.
 FT TRANSMEM 21 41
 FT TRANSMEM 143 163
 FT TRANSMEM 173 193
 FT TRANSMEM 198 218
 FT TRANSMEM 233 253
 FT TRANSMEM 258 278
 FT TRANSMEM 342 362
 FT TRANSMEM 374 394
 FT TRANSMEM 407 427
 FT TRANSMEM 431 451
 FT TRANSMEM 489 509
 FT CARBOHYD 57 57
 FT CARBOHYD 64 64
 FT CARBOHYD 91 91
 FT CONFLICT 275 275 W > G (IN REF. 2).
 SQ SEQUENCE 557 AA; 62567 MW; 172472E7B0A5F043 CRC64;

Query Match 76.1%; Score 2165; DB 1; Length 557;
 Best Local Similarity 73.2%; Pred. No. 1.5e-137;
 Matches 408; Conservative 68; Mismatches 75; Indels 6; Gaps 3;

QY 1 MRDYDEVTAFILGEGPQRLFFLFFLSASIPNGFNGSVVFLAGTPEHRCRVPDANLSS 60
 DB 1 MRDYDEVTAFILGEGPQRLFFLFFLSASIPNGFNGSVVFLAGTPEHRCRVPDANLSS 60
 QY 61 AWRNNSVPLRLRDGREGVPHSCSRRLATIANFSAIGLEPGRDVDLQLEQESCLDGEFS 120
 DB 61 AWRNHSIPLKTDGQVQPCRRYRLATIANFSAIGLEPGRDVDLQLEQENCLDGEYN 120
 QY 121 ODVYLSTVVTWNLVCEDNKVPVLTSLFFVGLGSEVSGQLSDRFGRKNVLPATMAVO 180
 DB 121 KDVFLSTVITVWDLVCKDDKAPLTTSLFFVGLGSEVSGQLSDRFGRKNVLPATMAVO 180
 QY 181 TGFSLQIFSISWEMFTVLFVIVGQISNYVAFILGTILGKSVRIIFSTLGVCTFFA 240
 DB 181 TGFSLQIFSISWEMFTVLFVIVGQISNYVAFILGTILGKSVRIIFSTLGVCTFFA 240
 QY 241 VGYMLLPFAFYFIRDWRMLLALTPVGLVPLVWFWIPESPRWLISQRRPREADITQKA 300
 DB 241 FGMVLPFAFYFIRDWRMLLALTPVGLVPLVWFWIPESPRWLISQRRPREADITQKA 300
 QY 301 AKMNTAVPAVED--SVEELNPLKQKAFITLDLFRTRNIAIMTMSLLMLLSVGYFA 358
 DB 301 AKFNIVAPSTIFDSELODLSKPKQSHHYYDLVTRNRIRIIMSLMLTISVGYFG 360
 QY 359 LSLDAPNLHGDAYLNCFLSALIEPAYITAWLLRLTPRRYIIAAVLFWGGVLLFTQLV 418
 DB 359 LSLDAPNLHGDAYLNCFLSALIEPAYITAWLLRLTPRRYIIAAVLFWGGVLLFTQLV 418
 QY 419 PVDYFSLGLVLMCKFOITSAFSMLYVFTAEIPTLVRNMAVGTSTASRVGSIIAPYF 478
 DB 421 PSELFTALVWVGKFGITSAYSVMVYVYTAELPTVVRNMGVGSVSTASRLGSILSPYF 480
 QY 479 VYLGAYNMLPYVWGLSVLVLIGITLFPESLGMTLPETLEOMQKVKWFRSGK---KTR 535
 DB 481 VYLGAYDRPLPYLWGLSVLVLITLFLPESFGTLPDIDQMLRVKGMKHKRTPSHTR 540
 QY 536 DSMETEENPKVL-ITAF 551
 DB 541 MLKDGQERTILKSTAF 557

RESULT 3
 OCN2_MOUSE
 ID OCN2_MOUSE STANDARD; PRT; 557 AA.
 AC Q920E8;

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U12965; AAA20607.1; -.
DR WormPep: F23F12.5; CE01252.
DR InterPro: IPR001066; -.
DR Pfam: PF00083; sugar_tr; 1.
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 24 44 POTENTIAL.
FT TRANSMEM 57 77 POTENTIAL.
FT TRANSMEM 102 122 POTENTIAL.
FT TRANSMEM 137 157 POTENTIAL.
FT TRANSMEM 200 220 POTENTIAL.
FT TRANSMEM 235 255 POTENTIAL.
FT TRANSMEM 320 340 POTENTIAL.
FT TRANSMEM 348 368 POTENTIAL.
FT TRANSMEM 379 399 POTENTIAL.
FT TRANSMEM 410 430 POTENTIAL.
FT TRANSMEM 432 452 POTENTIAL.
FT TRANSMEM 515 535 POTENTIAL.
FT TRANSMEM 547 567 POTENTIAL.
FT TRANSMEM 583 603 POTENTIAL.
FT TRANSMEM 614 634 POTENTIAL.
FT TRANSMEM 678 698 POTENTIAL.
SQ SEQUENCE 751 AA; 84832 MW; A6C4F43540295EFC CRC64;

Query Match 13.8%; Score 393; DB 1; Length 751;
Best Local Similarity 27.0%; Pred. No. 3.5e-19;
Matches 119; Conservative 91; Mismatches 184; Indels 46; Gaps 14;
QY 127 TVVTENLVCEDN-KWVPLTTSLEFVGLGSGVSDRGKRVLFATMAVQTGFSF 185
DB 303 SWVDQKFCGKAYDAWVAWTIQIIVLGITVYGHLDHFGKRPVSEFGVILGV 362
QY 186 LQIFSIWSEMTVLFVIVGMGOISVNVVAFILGTILGKSVRIIFSTLGVCTFFAVGY-- 243
DB 363 ASGFAPSWVEFAAFRIVGTSIASILIVFYAYILEFIEPEQRVELR-----SFFNNGYAR 417
QY 244 MLLPLPAYFIRDMRLLLATVPGVLCVPLWFWFIPESPRLISQRRFREAEDIIQAAKM 303
DB 418 LVFTLACFCIGYWRSAAIATSLSLPLPVLVLLPESPKWFTKRRFRDARAARVAVL 477
QY 304 NNTAVPAV-----IFDSVVELNPLKQKAFIL-DLFRTRNIAIMTMSLLWMLTSV 354
DB 478 --SGIPYVNDQOSIESEKLEE-----KTKIYTMKDLFTSWIATYRTIVGSLWFSTSL 531
QY 355 GYFALSLDAPNLHGDAYLNCFLSALIEIPAYITAMLLLTLP---RRYI-----IAAVLF 406
DB 532 SAFGSDLSNGLAGNLYLQFVSGAVTAFKIFVFLDITVPSFDRRLHQVQJAMILC 591
QY 407 WGGGVLLFTOLVPV-----DIYFLSIGLVMLGKFGITSAFSLMYVFTAEYPTLVR 457
DB 592 Y--CVIMVLMILPESDCGSGSRDLAIIINTIGVSFIBT--WDACYLVAVECEPPTKIR 647
QY 458 NMAVGVTSTASRGSIIAPFVYVVLGAYNRMPLVYVNGSL-TVLIGITLTFPPESLGMPL 516
DB 648 TIGIGTCSLLARTGALLAQMAVLSDIYRPAPYAVVCSIGTISLLISCVFLPDTKGVDL- 706
QY 517 ETLEQMKQKWRFSRGGKTRD 536
DB 707 AALDPTTELDYDRKKSMTEN 726
RESULT 5
YMP3_CAEEL STANDARD; PRT; 1222 AA.
ID YMP3_CAEEL

AC Q10947;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE HYPOTHETICAL 139.9 KDA PROTEIN B0361.3 IN CHROMOSOME III.
GN B0361.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RC SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2.
RA Du Z.;
RL Submitted (FEB-1994) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U00031; AAA50622.1; -.
DR WormPep: B0361.3; CE00752.
DR Pfam: PF00083; sugar_tr; 1.
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 51 71 POTENTIAL.
FT TRANSMEM 148 168 POTENTIAL.
FT TRANSMEM 182 202 POTENTIAL.
FT TRANSMEM 232 252 POTENTIAL.
FT TRANSMEM 263 283 POTENTIAL.
FT TRANSMEM 307 327 POTENTIAL.
FT TRANSMEM 329 349 POTENTIAL.
FT TRANSMEM 355 375 POTENTIAL.
FT TRANSMEM 382 402 POTENTIAL.
FT TRANSMEM 413 433 POTENTIAL.
FT TRANSMEM 442 462 POTENTIAL.
SQ SEQUENCE 1222 AA; 139868 MW; CBA42A80A254FB80 CRC64;
Query Match 11.2%; Score 317.5; DB 1; Length 1222;
Best Local Similarity 20.3%; Pred. No. 6.4e-14;
Matches 114; Conservative 106; Mismatches 219; Indels 123; Gaps 14;
QY 3 DYDEVIAFLGEGWPF-----QRLIFFLISAIIPNGFMGMSVWFLAGTPEHRCRVPD 54
DB 35 DPDRFVEAYGAYGKYQIFTVVLVQTLNFFYSSSMYI-----MSFVOL--NLEKQCE-- 83
QY 55 AANLSSAWRNNSVPLRLRDGVEPHSCSRVRLATIANFSALEPGRDVDLQLEQESCL 114
DB 84 -----YKNETIP-----ISETCO-----IETESSKAFGNLNGEYC- 113
QY 115 DGWFEFSODVYL-----STVVTENLVCEDNKMKVPLTTSLEFVGLGSGVSGQLSDR 166
DB 114 --GIAENTLVNVTNQKASTNLIVDFDLSCSHWFFQFEGLTIFTIGAVIAPVPMSLADR 170
QY 167 FGRKNVLFATMAVQTGFSFQIFSIWSEMTVLFVIVGMGOISVNVVAFILGTILGKSV 226
DB 171 YGRKPIIVTTAILAFLANMAASFSFNEAIFLILRAFIACGSDSYLSVASVATCEYLSEKA 230
QY 227 RIIFSTLGVCTFFAVGYMLLPLFAFYFIRDMRLLLATVPGVLCVPLWFWFIPESPRWLIS 286
DB 231 R-AMITVVYVNVAVSLGMVMTLLVLTMTDDMRVYFISLPGVGFALWYLFESPHMLIT 289
QY 287 QRRFREAEIIQKAAKMNNTAVPAVIEDSVBELNPLKQKAFILDLFRTRNIAIMTMSL 346
DB 290 KNKTEKLYKTKTANRM-----VISL 310
QY 347 LLWMLTSVGYFALSIDAPNLHGDAYLNCFLSALIEIPAYITAMLLLTLPRTYIAAVLF 406

```

Db 311 V-----YFAISFMSVLEGGDQVOAFLYSSLIETPAGLAVIPLMMKMRKMIWICLV 362
QY 407 WGG-----GVLLFQLPVVDYVYFLSIGLVMGKFGITSAFSMLYVFTAEYPTLVLRNMAVG 462
Db 363 FQTLALIGVTVFL-----DSYEFLKIVMLVAKVMATIIYVHPINATEQFPVSVSLCF 417
QY 463 VTSASRVSGITAPYVYVYGAYNMLPYIVMGSLTVLIGIFTLPFESLGMTLPETLEQM 522
Db 418 LNMIPQSMGIINSPYVYKHYVMSPNWPFVYVIALESFISATLAFMLHETKNKKLPDIESL 477
QY 523 -----QKVKVPRSGKKTRDSM 538
Db 478 SYPSTNDLSAYRRSKSSSSV 499

RESULT 6
YU13_CAEEL
ID YU13_CAEEL STANDARD; PRT; 435 AA.
AC Q10917;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE HYPOTHETICAL 48.6 KDA PROTEIN B0252.3 IN CHROMOSOME II.
GN B0252.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Du Z., Waterston R.;
RA Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U23453; AAC46757.1; -
CC WormPep: B0252.3; CE02419.
CC InterPro: IPR001066; -
CC Pfam: PF00083; sugar_tr; 1.
CC KW Hypothetical protein; Transmembrane.
CC FT TRANSMEM 55 75 POTENTIAL.
CC FT TRANSMEM 384 404 POTENTIAL.
CC SEQUENCE 435 AA; 48571 MW; 5F6160359FA1840B CRC64;

Query Match 10.8%; Score 307; DB 1; Length 435;
Best Local Similarity 23.58; Pred. No. 1.le-13;
Matches 94; Conservative 73; Mismatches 155; Indels 78; Gaps 8;

QY 128 VVTENLVCEQNMKVPVLTSTLFFGVGLGVSFVSGQLSDRFGRKNVLFATMAVQTFGSLQ 187
Db 91 VADEPDLTGDSWLAESTTTFYVGNMIGGMFIPPLADHYGLRVPVYATVLLMAVGGMIS 150
QY 188 IFSISWENFTLVFVVGQISNVVAPILTEILGKSVRIIFTLGVCTFVAVGYMLLP 247
Db 151 AFSTSIMMFCIMRNIHGFYTAAGLACVGLGYENTPLRLR-FFTSVYFVGMVWVAGACFLG 209
QY 248 LFAYFIQWRMLLALTPGVV-LCVPLWVFTPEPSRMLISORREAREEDIQK-AARKNN 305
Db 210 LLAYILPDRVLMFCISVPNFVALLIYVTPESLHFLVSSQQNEKLEAMLEKIRGPKGD 269
QY 306 TAPPAVIFDSVEELNPLK-----QQKAFILDLRTRNIAIMTISLLWMLTSSVGYFALS 361
Db 270 ISASDIVERDRDNGSSFKTLCREIKTEKTTLFQ----- 302

```

```

QY 362 DAPNLHGDAYLNCFLSALIEIPAYITAWLLRLTPRRYITAAVLFWGGVLLFQLVPVD 421
Db 303 -----DRYI-----YLFWIGLILLYEFGKPL- 324
QY 422 YVFSIGLVMGKFGITSAFSMLYVFTAEYPTLVLRNMAVGTVTASRVSGIIAPYFVL 481
Db 325 FEPCA-----HEFGRSSSLHFFSDPHEQIFPTDGRKNKICGFCETLSRFGGMLSPYLSHL 378
QY 482 GAYNRMLPYIVMGSLTVLIGIFTLPFESLGMTLPETLEQ 521
Db 379 TAVHALAPAITLSLIAVSGGLLTLLPETLNTKLPSTIAE 418

RESULT 7
YU13_CAEEL
ID YU13_CAEEL STANDARD; PRT; 529 AA.
AC P30636; Q21101;
DT 01-APR-1993 (Rel. 25, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE HYPOTHETICAL 58.3 KDA PROTEIN ZK637.1 IN CHROMOSOME III.
GN ZK637.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA MEDLINE-92168156; PubMed-1538779;
RA Sulston J., Du Z., Thomas K., Wilson R., Hillier L., Staden R.,
RA Halloran N., Green P., Thierry-Mieg J., Olu L., Dear S., Coulson A.,
RA Craxton M., Durbin R.K., Berks M., Metzstein M., Hawkins T.,
RA Ainscough R., Waterston R.;
RA Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLE).
CC -1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: Z11115; CAA77460.1; -
CC EMBL: Z22175; CAA77460.1; JOINED.
CC EMBL: Z22175; CAA80131.1; -
CC EMBL: Z11115; CAA80131.1; JOINED.
CC PIR: S15786; S15786.
CC WormPep: ZK637.1; CE066638.
CC InterPro: IPR001066; -
CC Pfam: PF00083; sugar_tr; 1.
CC PROSITE: PS00216; SUGAR_TRANSPORT_1; FALSE_NEG.
CC PROSITE: PS00217; SUGAR_TRANSPORT_2; FALSE_NEG.
CC KW Hypothetical protein; Duplication; Transmembrane; Transport.
CC FT TRANSMEM 86 106 POTENTIAL.
CC FT TRANSMEM 122 142 POTENTIAL.
CC FT TRANSMEM 158 178 POTENTIAL.
CC FT TRANSMEM 238 258 POTENTIAL.
CC FT TRANSMEM 320 340 POTENTIAL.
CC FT TRANSMEM 373 393 POTENTIAL.
CC FT TRANSMEM 411 431 POTENTIAL.
CC FT TRANSMEM 482 502 POTENTIAL.
CC SEQUENCE 529 AA; 58317 MW; 8D2FF4CBA15ECD2D CRC64;

```



```
Query Match          10.4%; Score 294.5; DB 1; Length 529;
Best Local Similarity 25.6%; Pred. No. 9e-13;
Matches 110; Conservative 78; Mismatches 150; Indels 91; Gaps 20;

QY 134 LVCDNDKV-----PLTTSFFGVILGSPVSGQLSDRFG-RKNVLFAF-----MAVQTG 182
DB 110 LACE-WGISVQOALVTTCTVFSQMLSTFWCKICDRGRRKGLFTSLVACIMGVISG 167
QY 183 FSLQIFSIWEMFTVLF-----VIVMGQISNVVAF--ILGTEILGKSVRIIFSTLGYC 236
DB 168 -----MSPHFYVLLFFRGLGFGIGGVPQSWTLVAFELPTAORAKCVLIES----- 214
QY 237 TEFAVGVMLLPLPAYFIRD---WRMLLALTVP-GVLCVPLWVFIPESPRWLISQRRFRE 292
DB 215 -FWAIGAVFALLAYFVWESFGMRMLFUSLLPLGIFAVASF--LPESARFDMASGHPER 272
QY 293 AEDIIQKAAMNNTAVP-AVIFDSVEELNPLKQOKAFIL--DLFRTRNIAIMTMSLLW 349
DB 273 ALETLQAAARNRVOLPTGRLVSTKAGSESRGDIANLLSPDLRKT-----TILLWCW 326
QY 350 MLTSVGVFALSADPNL-----HG-----DAYLNCFLSALIEIPAY 385
DB 327 AITAFSYGMVLETTVLFQSHDECHGFLSGNGTQMEVCOPLTRSDYFDLLSTTLAEFPGL 386
QY 386 ITAWLLL-----RTLPRRYIIAAVLEWGGVLLFIQLVPVDYFYFLSIGVLMGKFGITS 439
DB 387 IITVLLIIEWFGRRKTMALAYVAF-----IFTLLYFCLDRFTVTV-LIFVARAFISG 438
QY 440 AFSMLYVFTAEYPTLVRNMAVGVTSTASRVGSIIAPYFYVILGAYNMLPYIVMGSLTVL 499
DB 439 AFQCAVYVTEVPTLRAVGLGTCAMARIGAIVA-----SEKSLSPIGIGYTAAIL 492
QY 500 IGIPTLFFP 508
DB 493 GLIASLSIP 501

RESULT 8
YCEI_BACSU          STANDARD;          PRT;          400 AA.
AC O34691;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE HYPOTHETICAL METABOLITE TRANSPORT PROTEIN IN RAPJ-OPUAA INTERGENIC
DE REGION.
GN YCEI.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Kumano M., Tamakoshi A., Yamane K.;
RT "A 32 kb nucleotide sequence from the region of the lincomycin-
RT resistance gene (22-25 degree) of the Bacillus subtilis chromosome and
RT identification of the site of the lin-2 mutation."
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB000617; BAA22256.1; -
DR EMBL; Z99105; CAB12089.1; -
```

```
DR Subtilist; BG12773; yceI.
DR InterPro; IPR001066; -.
DR Pfam; PF00083; sugat_tr; 1.
DR PROSITE; PS00216; SUGAR_TRANSPORT_1; FALSE_NEG.
DR PROSITE; PS00217; SUGAR_TRANSPORT_2; 1.
KW Hypothetical protein; Transport; Transmembrane.
FT TRANSMEM 15 35 POTENTIAL..
FT TRANSMEM 50 70 POTENTIAL..
FT TRANSMEM 78 98 POTENTIAL..
FT TRANSMEM 99 119 POTENTIAL..
FT TRANSMEM 143 163 POTENTIAL..
FT TRANSMEM 166 186 POTENTIAL..
FT TRANSMEM 218 238 POTENTIAL..
FT TRANSMEM 254 274 POTENTIAL..
FT TRANSMEM 281 301 POTENTIAL..
FT TRANSMEM 305 325 POTENTIAL..
FT TRANSMEM 344 364 POTENTIAL..
FT TRANSMEM 371 391 POTENTIAL..
SQ SEQUENCE 400 AA; 43708 MW; EOAE0CEE5DD27395 CRC64;

Query Match          10.2%; Score 290; DB 1; Length 400;
Best Local Similarity 25.5%; Pred. No. 1.3e-12;
Matches 105; Conservative 68; Mismatches 165; Indels 74; Gaps 16;

QY 116 GHEP-SQDV-YLSTVV-----TENNLVCEP-NWKVPLTTSLFVGVLLGSPVSGQLSDRFG 168
DB 19 GFLFDAMDVGILSFITIAALHVEWNLSPPEMKW-----IGSVNSIGMAGAFLLGLADRIG 74
QY 169 RKNVLFAFMAVGTGFSFLQIFSIWEMFTVLFVIVMGQISNVVAFILGTEIL-----G 223
DB 75 RKVFIITLLCESIGSGISAFVTSLSAFLLIRFVIGMLGGLGELPVASTLVSEAVVPERG 134
QY 224 KSVRIIFTGLVCTFFAVGYMLLPLPAYFI---RDWRMLLALTVPVGVLCVPLWVFIPE 280
DB 135 RVIVLLES-----FWAVGWLAALISYFVTPSGWQAALLLTALTAFYALVLRSLDPS 188
QY 281 PRWLISQRRFEAEEDIIQKAAMNNTAVPAVIFDSVEELNPLKQOKAFILDLFRTRNIAI 340
DB 189 PKY-----ESLSAKRRSM-----WENKSV-----WARQYIR 215
QY 341 MTIMSLILLMLTSVGYFALSIDAPN---LHGDAYLNCF---LSALIEIPAYITAWLLR 393
DB 216 PTVMLSIVWFCVVFYSYGMFLWLPSPVLLKGFMSIQSEYVLLMTLAQLPGYFSAWLLIE 275
QY 394 TPLPRYIIAAVLEWGGVLLFIQLVPVDYFYFLSIGVLMGKFGITSFMSLYVFTAEYLP 453
DB 276 KAGRKWILVVLIGTAGSAYFFGTADSLSLITAG-VLLSFFNL-GAWGVLYAVTPEQYP 333
QY 454 TLVRNMAVGVTSTASRVGSIIAPYFYVILGAYNMLPYIVMGSLTVLIGIFTL 505
DB 334 TAIRATGSGTTAAFGRIIGGIFGLLVGTAAARHI-----SFSVIFSIFCI 378
```

```
RESULT 9
YAAU_ECOLI          STANDARD;          PRT;          443 AA.
AC P31679; P31578; P75628;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHETICAL METABOLITE TRANSPORT PROTEIN IN CARB-KEFC INTERGENIC
DE REGION (ORF65/66).
GN YAAU.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia
OC Escherichia
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=92334977; Pubmed=1630901;
RA Yura T., Mori H., Nagai H., Nagata T., Ishihama A., Fujita N.,
```



```
DR Pfam; PF00083; sugar_tr; 1.
DR PROSITE; PS00216; SUGAR_TRANSPORT_1; FALSE NEG.
DR PROSITE; PS00217; SUGAR_TRANSPORT_2; FALSE NEG.
KW Hypothetical protein; Transport; Transmembrane; Inner membrane.
FT TRANSMEM 26 46
FT TRANSMEM 61 81
FT TRANSMEM 91 111
FT TRANSMEM 113 133
FT TRANSMEM 154 174
FT TRANSMEM 182 202
FT TRANSMEM 272 292
FT TRANSMEM 302 322
FT TRANSMEM 330 350
FT TRANSMEM 352 372
FT TRANSMEM 400 420
FT TRANSMEM 421 441
SQ SEQUENCE 459 AA; 49602 MW; BC8AB53BCB8BD077 CRC64;

Query Match      8.6%; Score 245; DB 1; Length 459;
Best Local Similarity 23.4%; Pred. No. 1.6e-09;
Matches 97; Conservative 81; Mismatches 177; Indels 60; Gaps 17;

QY 134 LYCEDN-----WKVPLTTSLF-----FVGVLGGSFVSGQLSDRFGKKNVLEA 175
DB 34 LVCWSNAVGGLLLAQLKALGWTDNSTTATSAITTAGMFLGALVGGIIGDKTGRNFAIL 93
QY 176 TNAVOTGFSFQIPISEHMEFTVLFVVGAGQISYVVAFLIGTEIL-GK-----SVRII 229
DB 94 YEAIHIAKSVVGAAPSPNDFIACRFVGVGLGALLTFLFAGTEYMPGNRGTWSSRVS 153
QY 230 FS---TLGVCFFAVGVMLLPFAFYFRDWRMLLALTVPGLVCPPL-WWFIPESRWLI 285
DB 154 FIGNNSYPLCSLIANG--LTPLS-AEWNRRVQLLIPAILSLIATLALAWYFPFESRWLE 210
QY 286 SORRPREADIQK-----AARNNTAVPAVFDVSVEELNPLKQKAFILDLFTRNI--- 338
DB 211 SGRIOEAEKVMRSIEEGVIRQTKPLPPVVIADGKAPQAPVYSAALLTGVLKRVILGS 270
QY 339 AITMISLLMLLVSGVFALSALDAPNLHGDAYLNCFLSALIEIP-AYTAWLLRLTLPR 397
DB 271 CVLIAMNVQVTLINWLTPIETQGNLKSIVLNTW--SMGAPGFIETAMLVMDKIPR 328
QY 398 RVIIAAVFWGGVLLFLTQLPVDYFYS--IGLYMLKFKIYSAFMLY-----VFYA 449
DB 329 KFM-----GVGLILIAVGYIYSLQTSMLLTILIGFLIT--FVYMTVCYASAVYP 379
QY 450 ELYPLVRNMAVGTSTASRVGSIAPY--FVYLGAYNRMLPYIVMGSITVLIGI 502
DB 380 EIWPTEAKLRGSGUANAVGRISGIAAPYAVAVALLSSYGVTVGTFILLGAVSIIVAI 434

RESULT 15
KHT2_KLJOLA
ID KHT2_KLJOLA STANDARD; PRT; 566 AA.
AC P53387;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE HEXOSE TRANSPORTER 2.
GN KHT2.
OS Kluyveromyces lactis (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
OX NCBI_TaxID=28985;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAG;
RX MEDLINE=98028406; PubMed=9363776;
RA Weirich J., Goffrini P., Kuger P., Ferrero I., Breunig K.D.;
RT "Influence of mutations in hexose-transporter genes on glucose
RT repression in Kluyveromyces lactis.";
RL Eur. J. Biochem. 243:248-257(1997).
```

Search completed: August 16, 2001, 14:04:00
Job time: 302 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 16, 2001, 13:58:25 ; Search time 36.39 Seconds

(without alignments)
927.935 Million cell updates/sec

Title: US-09-521-195-3

Perfect score: 2883

Sequence: 1 MRDYDEVTAFLGEMGFQRL.....HTRMLKDGQRPRTILKSTAF 557

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :	A.Geneseq_0601.*
1:	/SID88/gcgdata/geneseq/geneseq/AA1980.DAT.*
2:	/SID88/gcgdata/geneseq/geneseq/AA1981.DAT.*
3:	/SID88/gcgdata/geneseq/geneseq/AA1982.DAT.*
4:	/SID88/gcgdata/geneseq/geneseq/AA1983.DAT.*
5:	/SID88/gcgdata/geneseq/geneseq/AA1984.DAT.*
6:	/SID88/gcgdata/geneseq/geneseq/AA1985.DAT.*
7:	/SID88/gcgdata/geneseq/geneseq/AA1986.DAT.*
8:	/SID88/gcgdata/geneseq/geneseq/AA1987.DAT.*
9:	/SID88/gcgdata/geneseq/geneseq/AA1988.DAT.*
10:	/SID88/gcgdata/geneseq/geneseq/AA1989.DAT.*
11:	/SID88/gcgdata/geneseq/geneseq/AA1990.DAT.*
12:	/SID88/gcgdata/geneseq/geneseq/AA1991.DAT.*
13:	/SID88/gcgdata/geneseq/geneseq/AA1992.DAT.*
14:	/SID88/gcgdata/geneseq/geneseq/AA1993.DAT.*
15:	/SID88/gcgdata/geneseq/geneseq/AA1994.DAT.*
16:	/SID88/gcgdata/geneseq/geneseq/AA1995.DAT.*
17:	/SID88/gcgdata/geneseq/geneseq/AA1996.DAT.*
18:	/SID88/gcgdata/geneseq/geneseq/AA1997.DAT.*
19:	/SID88/gcgdata/geneseq/geneseq/AA1998.DAT.*
20:	/SID88/gcgdata/geneseq/geneseq/AA1999.DAT.*
21:	/SID88/gcgdata/geneseq/geneseq/AA2000.DAT.*
22:	/SID88/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	2883	100.0	557	20	AA191650 A protein with cat
2	2883	100.0	557	21	AA191650 Human carnitine tr
3	2513	87.2	557	20	AA191652 A protein with cat
4	2513	87.2	557	21	AA191652 Mouse OCTN2 amino
5	2513	87.2	557	21	AA191652 Mouse carnitine tr
6	2295.5	79.6	564	21	AA191652 Mouse OCTN3 protei
7	2218	76.9	551	20	AA191649 A protein with cat
8	2148	74.5	553	20	AA191651 A protein with cat
9	2148	74.5	553	21	AA191651 Mouse OCTN1 amino
10	785.5	27.2	560	21	AA191651 Human ORFX ORF2802
11	735	25.5	554	19	AA191651 Human liver cell c

12	724.5	25.1	540	22	AA191650 Murine organic ani
13	719	24.9	535	21	AA191650 Rat liver anion tr
14	712	24.7	556	17	AA191650 Rat OCT-1 protein.
15	693	24.0	542	21	AA191650 Human cerebral org
16	690	23.9	548	21	AA191650 A human organic an
17	686.5	23.8	551	21	AA191650 A human organic an
18	673	23.3	561	18	AA191650 Human osteoclast t
19	660.5	22.9	607	21	AA191650 Hydropobic domain
20	656.5	22.8	607	21	AA191650 Rat cerebral organ
21	652.5	22.6	537	18	AA191650 Mouse osteoclast t
22	644	22.3	545	22	AA191650 Human organic anio
23	641	22.2	550	21	AA191650 Human organic anio
24	638.5	22.1	551	20	AA191650 Rat organic anion
25	628	21.8	563	20	AA191650 Human organic anio
26	607	21.1	550	22	AA191650 Arabidopsis thalia
27	433	15.0	480	21	AA191650 Arabidopsis thalia
28	433	15.0	483	21	AA191650 Arabidopsis thalia
29	380.5	13.2	515	21	AA191650 Arabidopsis thalia
30	368.5	12.8	521	21	AA191650 Arabidopsis thalia
31	357.5	12.4	548	21	AA191650 Human organic cati
32	344.5	11.9	397	21	AA191650 Arabidopsis thalia
33	332	11.5	439	22	AA191650 Corynebacterium gl
34	327.5	11.4	378	21	AA191650 Arabidopsis thalia
35	315.5	10.9	359	21	AA191650 Arabidopsis thalia
36	306.5	10.6	297	21	AA191650 Arabidopsis thalia
37	302.5	10.5	520	20	AA191650 Human organic cati
38	302.5	10.5	520	21	AA191650 Human saccharide-t
39	302.5	10.5	520	21	AA191650 Human secreted pro
40	284	9.9	360	21	AA191650 Cat flea HMT synap
41	282	9.7	339	21	AA191650 Amino acid sequenc
42	280.5	9.7	231	17	AA191650 Partial human OCT-
43	258	8.9	742	21	AA191650 Human secreted pro
44	251	8.7	530	21	AA191650 Cat flea HMT synap
45	234	8.1	494	20	AA191650 Glucose transporte

ALIGNMENTS

RESULT 1	AA191650	AA191650 standard; Protein: 557 AA.
ID	AA191650	AA191650 standard; Protein: 557 AA.
XX	AA191650	AA191650 standard; Protein: 557 AA.
AC	AA191650	AA191650 standard; Protein: 557 AA.
XX	AA191650	AA191650 standard; Protein: 557 AA.
DT	23-JUN-1999	(first entry)
XX	23-JUN-1999	(first entry)
DE	A protein with cation transporting activity.	
XX	A protein with cation transporting activity.	
KW	Organic cation transporter; OCT1; OCT2; drug development; fatty liver;	
KW	heart disease; cancer; anti-tumour drug; anticancer drug.	
XX	Homo sapiens.	
OS	Homo sapiens.	
XX	WO9913072-A1.	
PN	WO9913072-A1.	
XX	18-MAR-1999.	
PD	18-MAR-1999.	
XX	07-SEP-1998;	98WO-JP04009.
PF	07-SEP-1998;	98WO-JP04009.
XX	20-MAY-1998;	98JP-0156660.
PR	08-SEP-1997;	97JP-0260972.
XX	(CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.	
PA	Nezu J, Oku A;	
XX	WPI; 1999-215062/18.	
XX	N-PSDB; AAX26880.	
DR	Genes homologous with organic cation transporters OCT1 and OCT2,	
DR	useful in design of new drugs for treatment of diseases due to	
PT	abnormality of the transporter functions	

QY	541	MLKDGQERPTILKSTAF 557	
Db	541	mlkdqgerptilkstaf 557	
RESULT 3			
AA	Y01652	standard; Protein; 557 AA.	
AC	Y01652;		
DT	23-JUN-1999	(first entry)	
XX	A protein with cation transporting activity.		
XX	Organic cation transporter; OCT1; OCT2; drug development; fatty liver;		
KW	heart disease; cancer; anti-tumour drug; anticancer drug.		
XX			
OS	Mus musculus.		
XX			
PN	WO9913072-A1.		
XX	18-MAR-1999.		
XX	07-SEP-1998; 98WO-JP04009.		
XX	20-MAY-1998; 98JP-0156660.		
PR	08-SEP-1997; 97JP-0260972.		
XX	(CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.		
XX			
PI	Nezu J, Oku A;		
XX	WPI; 1999-215062/18.		
DR	N-PSDB; AAX26902.		
XX	Genes homologous with organic cation transporters OCT1 and OCT2,		
PT	useful in design of new drugs for treatment of diseases due to		
PT	abnormality of the transporter functions		
XX			
PS	Claim 1; Page 75-79; 97pp; Japanese.		
XX	The present sequence represents a protein with cation transporting		
CC	activity. The genes are significantly homologous with organic cation		
CC	transporters OCT1 and OCT2. The genes may be used in drug development,		
CC	particularly in the treatment of diseases due to abnormality of the		
CC	organic cation transporter functions e.g. fatty liver, heart diseases		
CC	and cancers, by controlling such as by inhibition or activation.		
CC	Administration of anti-tumour and anticancer drugs in combination with		
CC	a transporter protein inhibiting agent allows the agents to penetrate		
CC	into the diseased cells to enhance the drug action.		
XX			
SQ	Sequence 557 AA;		
Query Match 87.2%; Score 2513; DB 20; Length 557;			
Best Local Similarity 85.5%; Pred. No. 1.9e-265;			
Matches 476; Conservative 39; Mismatches 42; Indels 0; Gaps 0;			
QY	1	MRDYDEVTAFLGEGPQRLIFLLSASIIIPNGFTGLSSVFLIATPEHRCRVPDAANLSS 60	
Db	1	mrdydevtaflgwgppqrlifllsasllpngftglssvflatpchrclvphtvnlss 60	
QY	61	AWRNHTVPLRLDRGVPVPHSCRRYRLATIANFSALEPGRDVDLQGLESCLDGWEFS 120	
Db	61	awrnhsipletkdgrvppqkrryrlatianselglepgrdvldleqescldgweyd 120	
QY	121	QDVLSTIVTEWNLVCDWDKAPLTISLFFVGVLLGSFISGQLSDRFGKKNVLFVTMGMO 180	
Db	121	kdvflstivtewldvckdwkapltslffvgvlmgsgfsgqlsdrfgkknvlfmgmq 180	
QY	181	TGFSFLQIFSKNFEMFVFLVGVGMQGISNYAAVFLGTTEILGKSVRIIFSTLIGVCIFYA 240	
Db	181	tgfsflqvfsnfmftvlfvlgmgqgisnyaaavflgtteilsksrllfatlgvcifya 240	
QY	241	FGYMWLPFLFAYFTRDWRMLLVALTMFGVLCVALWFIPESPRWLISQGRFEAEVIRKA 300	
Db	241	fgmvlplfayfirdwrmlllaltvpqvlcgaiwffipesprwlisggrikeaeavirka 300	
QY	301	AKANGIVVPSTIFDPSLODLSSKQOOSHNLDLRLTRNIRMTIMSILWMTISVGYFG 360	
Db	301	akingivapstifdpselqdnstkpqlhnydlirnrlnrvitmsilwltisvgyfg 360	
QY	361	LSLDTPNLRHGDIFVNCFLSAMVEPAYVLAWLLOLQVLPFRYSMATALLFGSGVLLPMQLV 420	
Db	361	lsldtpnlhgdlyvncfllaavepayvlawllqvlpryysaailfgsgvllmqvlv 420	
QY	421	PPDIYYIATVLMVWGKFCVTAAFSMVYVYVTAELYPVVRNMGVGVSTASRLSILSPYF 480	
Db	421	pselfyistalvmvgkfgitsaysmvvyvtaelyptvvrnmvgvstasrlsilsy 480	
QY	481	VYLGAYDRFLPYILMGSITLTALTFLFESFGTLPDIDQMLRVKGMKHKRTPSHTR 540	
Db	481	vylgaydrflpyilmgstltaltlffesfgvpldtidqmlrvkqikqwgqsgtr 540	
QY	541	MLKDGQERPTILKSTAF 557	
Db	541	mqkdgeesptvlkstaf 557	
RESULT 4			
AA	B20580	standard; Protein; 557 AA.	
AC	B20580;		
XX	11-DEC-2000 (first entry)		
DE	Mouse OCTN2 amino acid sequence.		
XX	Mouse; transporter; OCTN1; OCTN2; OCTN3; organic cation transporter;		
KW	identification; regulator; carnitine transport.		
XX	Mus musculus.		
XX	WO200046368-A1.		
PD	10-AUG-2000.		
XX	04-FEB-2000; 2000WO-JP00619.		
PR	05-FEB-1999; 99JP-0028406.		
XX	(CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.		
XX	Nezu J, Ose A;		
XX	WPI; 2000-586982/55.		
XX	Organic cation transporter gene OCTN3 expressed in testis for		
PT	identification of regulators of carnitine transport for use as drugs		
XX	Example 2; Fig 2; 58pp; Japanese.		
PS	The present invention describes a mouse organic cation transporter		
CC	protein (OCTN3). Also described are: (1) a method for screening		
CC	compounds for their ability to regulate the transport of an organic		
CC	cation into the cell, by generating a cell expressing OCTN3 at the		
CC	cell membrane, contacting with the compound and organic cation, and		
CC	observing the degree of transport of the organic cation; and (2) a		
CC	method for screening compounds for their ability to be transported into		
CC	the cell by OCTN3, by generating a cell expressing OCTN3 at the cell		
CC	membrane, contacting with the compound and observing the degree of		
CC	transport of the compound. OCTN3 can be used for the identification of		
CC	regulators of the transport of organic cations (especially carnitine)		

CC into cells by OCTN3, for use as drugs. The present sequence represents
 CC the mouse OCTN2 amino acid sequence, which is used in an example from
 CC the present invention.

XX Sequence 557 AA;

Query Match 87.2%; Score 2513; DB 21; Length 557;
 Best Local Similarity 85.5%; Pred. No. 1.9e-265;
 Matches 476; Conservative 39; Mismatches 42; Indels 0; Gaps 0;

QY 1 MRDYDEVTAFELGNGPQRIFFLLSASIIIPNGFTGLSSVFLIATPEHRCRVPDAANLSS 60
 Db 1 mrdydevtafelgngpqrifllsasiipngftglssvfliatpehrcrclvphntvnls 60
 QY 61 AWRNHTVPLRLDGRVPHSCRRYRLATIANFSALGLEPGRDVLQLESCLDGWEFS 120
 Db 61 awrnhsipletkdgrrvpqkrryrlatianselglepgrdvldleqesclgdgweyd 120
 QY 121 QDVYLSITVTWNLVCEDDWKAPLTISLFFVGLGSGFISGQSLDRGRKNVLFVTMGMO 180
 Db 121 kvdlstivtwedlvckddwkaptltslffvglmgsfsgqlsdrgrknvlfclmgmq 180
 QY 181 TGFSEFLQIFSKNFMFVFLVVGMOISNYAAAFVLGTEILGKSVRIIFSTLGVCIIFYA 240
 Db 181 tgfsflqvsvnfemftvflvvgmqisnyaaafvlgtelilsksrilifatlvgvcifya 240
 QY 241 FGVMVPLFAFYFIRDMRLVLTMPGVLCVLAWVFPESRWLISQGRFEAEVIRKA 300
 Db 241 fgvmvplfayfirdmrlvltmpgvlcgalwfpesprwlisqgrfkaevirka 300
 QY 301 AKANGIVVPSTIFDPSELQDLSSKKQOSHNLDLRLTNIRMTVIMSIMLWMTISVGYFG 360
 Db 301 akingivapstifdpseqlqdlstkpqlhhydlrtnrnirvmtimsilwltisvgyfg 360
 QY 361 LSLDTPNLHGDIFVNCFLSAMVEVPAYVLAWLLQYLPRYSMATALFLGGSVLLFMQLV 420
 Db 361 lsldtpnlhgdifvncflsamvevpayvawlllqylprysaaltflggsvllfmqlv 420
 QY 421 PPDLXYLATVLMVGRFGVTAAFSMYVYVTAELYPTVVRNMGVSGVSTASRLGSLSPYF 480
 Db 421 pselfylstalmvgrfgvtaafsmvyyvtaelyptvvrnmvgvsgvstasrlgsilspyf 480
 QY 481 VYLGAYDRFLPYILMGSLLILTAITLFLPESFGTLPDIDQMLRVKGMKHKRTPSHTR 540
 Db 481 vylygaydrflpyilmgslliltailtflpfesfgvplpdtidqmlrvkgmkhkrtpshtr 540
 QY 541 MLKDGQERPTILKSTAF 557
 Db 541 mldgqerptilkstaf 557

RESULT 5

AY83930

ID AY83930 standard; Protein; 557 AA.

XX AC AY83930;

XX DT 05-JUL-2000 (first entry)

XX DE Mouse carnitine transporter protein OCTN2.

XX KW Organic cation transportation; mouse; carnitine transporter protein;
 KW OCTN2; diagnosis; systemic carnitine deficiency; mutation; gene therapy;
 KW juvenile visceral steatosis.

XX OS Mus musculus.

XX PN WO200014210-A1.

XX XX 16-MAR-2000.

XX PF 07-SEP-1999; 99WO-JP04853.

XX 07-SEP-1998; 99JP-0252683.

XX (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.

XX Nezu J, Oku A;

XX WPI; 2000-256966/22.

XX N-PSDB; AAA09890.

XX Systemic carnitine deficiency gene OCTN2 encoding part of organic
 cation transporter, useful as diagnostic tool

XX Example 1; Page 47-51; 106pp; Japanese.

XX This sequence represents the mouse carnitine transporter protein OCTN2.
 CC The coding sequence of the corresponding human protein can be used as a
 CC target for diagnosis of systemic carnitine deficiency by detecting the
 CC presence of mutations in the sequence, especially seen in the disease
 CC juvenile visceral steatosis (jvs). The wild type OCTN2 gene can be used
 CC in the gene therapy of the disease state.

XX Sequence 557 AA;

Query Match 87.2%; Score 2513; DB 21; Length 557;
 Best Local Similarity 85.5%; Pred. No. 1.9e-265;
 Matches 476; Conservative 39; Mismatches 42; Indels 0; Gaps 0;

QY 1 MRDYDEVTAFELGNGPQRIFFLLSASIIIPNGFTGLSSVFLIATPEHRCRVPDAANLSS 60
 Db 1 mrdydevtafelgngpqrifllsasiipngftglssvfliatpehrcrclvphntvnls 60
 QY 61 AWRNHTVPLRLDGRVPHSCRRYRLATIANFSALGLEPGRDVLQLESCLDGWEFS 120
 Db 61 awrnhsipletkdgrrvpqkrryrlatianselglepgrdvldleqesclgdgweyd 120
 QY 121 QDVYLSITVTWNLVCEDDWKAPLTISLFFVGLGSGFISGQSLDRGRKNVLFVTMGMO 180
 Db 121 kvdlstivtwedlvckddwkaptltslffvglmgsfsgqlsdrgrknvlfclmgmq 180
 QY 181 TGFSEFLQIFSKNFMFVFLVVGMOISNYAAAFVLGTEILGKSVRIIFSTLGVCIIFYA 240
 Db 181 tgfsflqvsvnfemftvflvvgmqisnyaaafvlgtelilsksrilifatlvgvcifya 240
 QY 241 FGVMVPLFAFYFIRDMRLVLTMPGVLCVLAWVFPESRWLISQGRFEAEVIRKA 300
 Db 241 fgvmvplfayfirdmrlvltmpgvlcgalwfpesprwlisqgrfkaevirka 300
 QY 301 AKANGIVVPSTIFDPSELQDLSSKKQOSHNLDLRLTNIRMTVIMSIMLWMTISVGYFG 360
 Db 301 akingivapstifdpseqlqdlstkpqlhhydlrtnrnirvmtimsilwltisvgyfg 360
 QY 361 LSLDTPNLHGDIFVNCFLSAMVEVPAYVLAWLLQYLPRYSMATALFLGGSVLLFMQLV 420
 Db 361 lsldtpnlhgdifvncflsamvevpayvawlllqylprysaaltflggsvllfmqlv 420
 QY 421 PPDLXYLATVLMVGRFGVTAAFSMYVYVTAELYPTVVRNMGVSGVSTASRLGSLSPYF 480
 Db 421 pselfylstalmvgrfgvtaafsmvyyvtaelyptvvrnmvgvsgvstasrlgsilspyf 480
 QY 481 VYLGAYDRFLPYILMGSLLILTAITLFLPESFGTLPDIDQMLRVKGMKHKRTPSHTR 540
 Db 481 vylygaydrflpyilmgslliltailtflpfesfgvplpdtidqmlrvkgmkhkrtpshtr 540
 QY 541 MLKDGQERPTILKSTAF 557
 Db 541 mldgqerptilkstaf 557

RESULT 6

AAB20578

ID AAB20578 standard; Protein; 564 AA.

XX AAB20578;
 AC 11-DEC-2000 (first entry)
 DT Mouse OCTN3 protein SEQ ID NO:1.
 XX Mouse; transporter; OCTN1; OCTN2; OCTN3; organic cation transporter;
 DE identification; regulator; carnitine transport.
 XX Mus musculus.
 OS WO2000046368-A1.
 PN 10-AUG-2000.
 PD 04-FEB-2000; 2000WO-JP00619.
 XX 05-FEB-1999; 99JP-0028406.
 PR (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
 PA Nezu J, Ose A;
 XX WPI; 2000-586982/55.
 DR N-PSDB; AAA88053.
 XX Organic cation transporter gene OCTN3 expressed in testis for
 PT identification of regulators of carnitine transport for use as drugs
 XX Claim 1; Page 34-39; 58pp; Japanese.
 CC The present invention describes a mouse organic cation transporter
 CC protein (OCTN3). Also described are: (1) a method for screening
 CC compounds for their ability to regulate the transport of an organic
 CC cation into the cell, by generating a cell expressing OCTN3 at the
 CC cell membrane, contacting with the compound and organic cation, and
 CC observing the degree of transport of the organic cation; and (2) a
 CC method for screening compounds for their ability to be transported into
 CC the cell by OCTN3, by generating a cell expressing OCTN3 at the cell
 CC membrane, contacting with the compound and observing the degree of
 CC transport of the compound. OCTN3 can be used for the identification of
 CC regulators of the transport of organic cations (especially carnitine)
 CC into cells by OCTN3, for use as drugs. The present sequence represents
 CC mouse OCTN3.
 XX Sequence 564 AA;

Query Match 79.6%; Score 2295.5; DB 21; Length 564;
 Best Local Similarity 80.6%; Pred. No. 1.1e-241;
 Matches 435; Conservative 54; Mismatches 48; Indels 3; Gaps 1;
 1 MRDYDEVTAFLGEGWGPFORLIFLLSASIPNGFTGLSVFLIATPEHRCRVPDAANLSS 60
 1 mldydevtaflgwggtfqrllflsasiipngftglsvflaipaehrcrripdvtvlss 60
 61 AWRNHTVPLRGDREVPCHSRRLATIANFSALEGRDVLGLQESCLDGWEFS 120
 61 awrnhsipmetkdgpevpqckryrlatiansfalegrdvdldleqencldgweyd 120
 121 QDVYLSITVTENLVCEDDWKAPLTSLFFVGLVGSFISGQLSDRFGRKNVLFVTMGWQ 180
 121 kdvlstivtwedlvckddwkapltslffvglvgsfsgqlsdrlfgrknvlftmghm 180
 181 TGFSEFLQIFSKNFEMFVFLVGLVGMQISNYAAAFVLGTSVRIIFSTLGVCFIYA 240
 181 tgfssiqvsvnfemflltytlvgmghisnyaaafvlgtemlsksvriifatlvgciffa 240
 241 FGWVPLPLFAYFIRDMWMLLVALTMPGVLVLCVAMWFIPESPRWLISQGRFEAEVIIRKA 300
 241 fgfmvplplfayfirewrrillaitlpgvlcalwvfipesprwllsggrikeaeavlrka 300

QY 301 AKANGIVVPSTIFDPSE---LQDSSKKQOSHNTDLLRTNIRMTVMTIMSLMMTISVG 357
 DB 301 akingivapstifopsetnklqddsskpkqshhlydrtvtpnrlitimsiilwltisvg 360
 QY 358 YFGLSLDTPNLHGDI FVNCFLSAMVEVPAYVLAWLLEQLYLPFRYSMATALFLGGSVLLFM 417
 DB 361 yfglsldtpnlnyincflaavevpayvlawlllqhvsrrysmagsiflgsgvllv 420
 QY 418 QLVPPDLYLATVLMVGKFGVTAFAFMSVYVYAEIYPTVVRNMGVSVSTASRLGSILS 477
 DB 421 qlvpsdlylatvlmvgkfgitsaysmvvyvtaelyptvvrnmgvsvstasrlgsils 480
 QY 478 PYFVYLGAIDRELPYILMGSLTILTAITLFLPESFCTPLDITDQMLRVKGMKHKRTPS 537
 DB 481 pyfvyldaydrilpymgsitlitaiffpessgvsipetidemqvkxkikqrgsls 540
 RESULT 7
 AAY01649
 ID AAY01649 standard; Protein; 551 AA.
 XX
 AC AAY01649;
 XX
 DT 23-JUN-1999 (first entry)
 XX
 DE A protein with cation transporting activity.
 XX Organic cation transporter; OCT1; OCT2; drug development; fatty liver;
 KW heart disease; cancer; anti-tumour drug; anticancer drug.
 KW
 OS Homo sapiens.
 XX
 PN W09913072-A1.
 XX
 PD 18-MAR-1999.
 XX
 PF 07-SEP-1998; 98WO-JP04009.
 XX
 PR 20-MAY-1998; 98JP-0156660.
 PR 08-SEP-1997; 97JP-0260972.
 XX
 PA (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
 XX
 PI Nezu J, Oku A;
 XX
 DR WPI; 1999-215062/18.
 DR N-PSDB; AAX26879.
 XX
 PT Genes homologous with organic cation transporters OCT1 and OCT2,
 PT useful in design of new drugs for treatment of diseases due to
 PT abnormality of the transporter functions
 XX
 PS Claim 1; Page 41-45; 97pp; Japanese.
 XX
 CC The present sequence represents a protein with cation transporting
 CC activity. The genes are significantly homologous with organic cation
 CC transporters OCT1 and OCT2. The genes may be used in drug development,
 CC particularly in the treatment of diseases due to abnormality of the
 CC organic cation transporter functions e.g. fatty liver, heart diseases
 CC and cancers, by controlling such as by inhibition or activation.
 CC Administration of anti-tumour and anticancer drugs in combination with
 CC a transporter protein inhibiting agent allows the agents to penetrate
 CC into the diseased cells to enhance the drug action.
 XX
 SQ Sequence 551 AA;

Query Match 76.9%; Score 2218; DB 20; Length 551;
 Best Local Similarity 75.9%; Pred. No. 3.1e-233;
 Matches 423; Conservative 57; Mismatches 71; Indels 6; Gaps 3;
 1 MRDYDEVTAFLGEGWGPFORLIFLLSASIPNGFTGLSVFLIATPEHRCRVPDAANLSS 60
 1 mldydevtaflgwggtfqrllflsasiipngftglsvflaipaehrcrripdvtvlss 60

Db 1 mrdydeviaflegwpgfqrlliflllsaslipngfngmsvflagtphehrcrvpdaanlss 60
 Qy 61 AWRNHTVPLRLRDGREGVPHSCRRYRLATIANFSALGLEPGRDVLGQLEQESCLDGEWFS 120
 Db 61 awrnsvplrlrdgrevphscsryrlatiansaiglegrdvldglegescldgews 120
 Qy 121 QDYLSTIVTENLVCEDDWKAPLTISLFFVGVLLGSFISGQLSDRFRGKNVLFVTMGQ 180
 Db 121 qdyistvtenlvceedwnkvphttslffvgvllgsfvsqglsdrfrgknvlfatmaq 180
 Qy 181 TGFSLQIFSKNFEMFVFLVGVGMGOISNYVAAPVLGTEILGKSVRIIFSTLGVCIYA 240
 Db 181 tgfslqifsknfemfvlvfgvngmsvnyvafilgtellgksvriifstlgtvctffa 240
 Qy 241 FGMYLPLFAYFIRDRWMLLVALTMGPVLCVALMWFIPESPRWLISQGRFEEAEVIIRKA 300
 Db 241 vgymlplfayfirdwrmllaltvgvclvplwlfipesprwlisqrfrfreaedliqka 300
 Qy 301 AKANGIVPSTTFDSELDLSSKKQSHNLDLRTWIRNMTIMSLMWTISVGYFG 360
 Db 301 akmntavpavifd--sveelnplkqgkafildiftrniatmtimsllwmltsvgyfa 358
 Qy 361 LSLDTPNLHGDIFVNCFLSAMVEVPAYVLAWLLLOYLPRRYSMATALFLGGSVLLFMOLV 420
 Db 359 lsldapnlhgdvlnclfsaliespaytawllrtlpriiaavlfwggvllfiqlv 418
 Qy 421 PPDLYLATVLMVKGKFGVTAAFSMVYVYTAELPTVVRNMGVSVSTASRLGSILSPYF 480
 Db 419 pvyvylfslglvmlgkfgitsafsmlyvftaelptlvrmnmgvststasrvgsliapyf 478
 Qy 481 VYLGYDRFLPVLGSLTILTAITLFLPESFGTLPDITDQMLRVKGMKHKRTPSHTR 540
 Db 479 vylgaydrflpvlmgsltlvlgiflfpeslgnltlpetlegmqkvkfrsgk---ktr 535
 Qy 541 MLKDGQERPTILKSTAF 557
 Db 536 dsmeteepkvl-itaf 551

RESULT 8
 ID AAY01651
 XX AAY01651 standard; Protein; 553 AA.
 AC AAY01651;
 XX
 DT 23-JUN-1999 (first entry)
 XX
 DE A protein with cation transporting activity.
 XX
 KW Organic cation transporter; OCT1; OCT2; drug development; fatty liver;
 KW heart disease; cancer; anti-tumour drug; anticancer drug.
 XX
 OS Mus musculus.
 XX
 PN W09913072-A1.
 XX
 PD 18-MAR-1999.
 XX
 PF 07-SEP-1998; 98W0-JP04009.
 XX
 PR 20-MAY-1998; 98JP-0156660.
 PR 08-SEP-1997; 97JP-0260972.
 XX
 XX (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
 XX
 PI Nezu J, Oku A;
 XX
 DR WPI; 1999-215062/18.
 DR N-PSDB; AAX26898.
 XX
 PT Genes homologous with organic cation transporters OCT1 and OCT2,
 PT useful in design of new drugs for treatment of diseases due to
 PT abnormality of the transporter functions

XX
 PS
 CC The present sequence represents a protein with cation transporting
 CC activity. The genes are significantly homologous with organic cation
 CC transporters OCT1 and OCT2. The genes may be used in drug development,
 CC particularly in the treatment of diseases due to abnormality of the
 CC organic cation transporter functions e.g. fatty liver, heart diseases
 CC and cancers, by controlling such as by inhibition or activation.
 CC Administration of anti-tumour and anticancer drugs in combination with
 CC a transporter protein inhibiting agent allows the agents to penetrate
 CC into the diseased cells to enhance the drug action.
 XX
 SQ Sequence 553 AA;

Query Match 74.5%; Score 2148; DB 20; Length 553;
 Best Local Similarity 72.2%; Pred. No. 1.4e-225;
 Matches 402; Conservative 67; Mismatches 84; Indels 4; Gaps 2;
 Qy 1 MRDYDEVIAFLEGWPGFQRLIFLLSASLIPNGFTGLSSVFLIATPEHRCRVDAANLSS 60
 Db 1 mrdydeviaflegwpgfqrlliflllsaslipngfngmsvflagtphehrcrvpdtvnls 60
 Qy 61 AWRNHTVPLRLDGRVPHSCRRYRLATIANFSALGLEPGRDVLGQLEQESCLDGEWFS 120
 Db 61 swrnhsipltkdgrvpgqscrryrlatiansamlepgqdvldleqesclldgweyd 120
 Qy 121 QDYLSTIVTENLVCEDDWKAPLTISLFFVGVLLGSFISGQLSDRFRGKNVLFVTMGQ 180
 Db 121 kdlfstivtenlvceedwktpltslffvgvllgsfvsqglsdrfrgkvlfatmaq 180
 Qy 181 TGFSLQIFSKNFEMFVFLVGVGMGOISNYVAAPVLGTEILGKSVRIIFSTLGVCIYA 240
 Db 181 tgfslqifstnwemfvlfaivngmsvnyvafilgtellsksvriifstlgtvctffa 240
 Qy 241 FGMYLPLFAYFIRDRWMLLVALTMGPVLCVALMWFIPESPRWLISQGRFEEAEVIIRKA 300
 Db 241 igymvplfayfirdwrmllaltlpglfcvplwlfipesprwlisgrfreaedliqka 300
 Qy 301 AKANGIVPSTTFDSELDLSSKKQSHNLDLRTWIRNMTIMSLMWTISVGYFG 360
 Db 301 akmslvapagifdplelqelskqkvliildftrniatitvmaavlmltsvgyfa 360
 Qy 361 LSLDTPNLHGDIFVNCFLSAMVEVPAYVLAWLLLOYLPRRYSMATALFLGGSVLLFMOLV 420
 Db 361 lslnvplnhgdvylncflslgievpayftawllrtlpriiaavlfwggvllfiqvv 420
 Qy 421 PPDLYLATVLMVKGKFGVTAAFSMVYVYTAELPTVVRNMGVSVSTASRLGSILSPYF 480
 Db 421 pedyntvsglvmgkfgitsafsmlyvftaelptlvrmnmgvstasrvgsliapyf 480
 Qy 481 VYLGYDRFLPVLGSLTILTAITLFLPESFGTLPDITDQMLRVKGMKHKRTPSHTR 540
 Db 481 vylgaydrflpvlmgsltlvlgiflfpesfgtvlpenlegmqkvrgfcgk---kat 537
 Qy 541 MLKDGQERPTILKSTAF 557
 Db 538 vsvdreespkvl-itaf 553

RESULT 9
 ID AAB20579
 XX AAB20579 standard; Protein; 553 AA.
 AC AAB20579;
 XX
 DT 11-DEC-2000 (first entry)
 XX
 DE Mouse OCTN1 amino acid sequence.
 XX
 KW Mouse; transporter; OCTN1; OCTN2; OCTN3; organic cation transporter;
 KW identification; regulator; carnitine transport.

OS Mus musculus.
 PN WO200046368-A1.
 XX 10-AUG-2000.
 XX 04-FEB-2000; 2000WO-JP00619.
 XX 05-FEB-1999; 99JP-0028406.
 XX (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
 XX Nezu J, Ose A;
 XX WPI; 2000-586982/55.
 XX Organic cation transporter gene OCTN3 expressed in testis for
 PT identification of regulators of carnitine transport for use as drugs
 XX Example 2; Fig 2; 58pp; Japanese.
 XX The present invention describes a mouse organic cation transporter
 CC protein (OCTN3). Also described are: (1) a method for screening
 CC compounds for their ability to regulate the transport of an organic
 CC cation into the cell, by generating a cell expressing OCTN3 at the
 CC cell membrane, contacting with the compound and organic cation, and
 CC observing the degree of transport of the organic cation; and (2) a
 CC method for screening compounds for their ability to be transported into
 CC the cell by OCTN3, by generating a cell expressing OCTN3 at the cell
 CC membrane, contacting with the compound and observing the degree of
 CC transport of the compound. OCTN3 can be used for the identification of
 CC regulators of the transport of organic cations (especially carnitine)
 CC into cells by OCTN3, for use as drugs. The present sequence represents
 CC the mouse OCTN1 amino acid sequence, which is used in an example from
 CC the present invention.
 XX Sequence 553 AA;
 SQ
 Query Match 74.5%; Score 2148; DB 21; Length 553;
 Best Local Similarity 72.2%; Pred. No. 1.4e-225;
 Matches 402; Conservative 67; Mismatches 84; Indels 4; Gaps 2;
 QY 1 MRDYDEVTAFLEWGPFQRLIFLLSASIPNGFTGLSSVFLIATPEHRCRVPDANLSS 60
 Db 1 mrdydeviafigewgpfqrlifllsasipngfmgsvvfiagtpchrclypdtvniss 60
 QY 61 AWRNHTVPLRLDGRVPHSCRRYRLATIANFSALGLEPGRDVLGQLQESCLDGWERS 120
 Db 61 swrnhsiptekdgrvpgscrryrlatiansamlegpgqdvdlqeqescldgweyd 120
 QY 121 QDYLSTIVTEWNLVCEDDWKAPLTLISLRFVGLLSGFTSGQLSDRFGKKNLFTVTMGQ 180
 Db 121 kdiflstivtewnlvceddwktltslrfvglcgsfgqlsdrfgrkklvatmavq 180
 QY 181 TGFSELOIFSKNEFMFVFLVINGMGQISNYAAFLVGLTEILGKSVRIIFSLGVCIFYA 240
 Db 181 tgrsfvqlifstnwemftvflvngmqisnyvavfllgteilsksvriifstlgtctffa 240
 QY 241 FGVMVLPFLFAFYTRDWRMLVALTMPGVLCLWALWTFIPSPRWLISQGRFEAEIIRKA 300
 Db 241 igrvmvlpflfayfirdwrmlvaltlp9lfcvplwlfipspwrlisqrfaeeqiika 300
 QY 301 AKANGIVVPTIFDPSELQSLSSKKQOSHNDLLRTNIRMTIMSIMLWMTISVGYGF 360
 Db 301 akmsivapagifdpelqelnsikqkvilldlfrtniativmavmlwmltsvgyfa 360
 QY 361 LSLDTPNLHGDIFVNCFLSMEVPAVLANLLQLVLPYRYSMATALFLGGSVLLPMQLV 420
 Db 361 lslnvpnlhgdvynclflsiegvepavftawllrltprlyiaigvlfvggvllliqv 420
 QY 421 PPDLYLATVLMVGKFGVTAAFSMVYVYTAELYPTVWRNMGVGVSSASRLGSLIPYF 480

Db 421 pedynfvsiglmglkgfitaefsmlyvftaeelyptlvrmavgitsmasrvgsilapyf 480
 QY 481 VYLGAIRFLPYILMGSLITLITLFLPESFCTPLPDTIDQMLRVKGMKHKRTPSHTR 540
 Db 481 vylgaynrllpyilmgslitlilitflpessfgvtlpenlegmqkvrgrcgk---kst 537
 QY 541 MLKDGQERPTILKSTAF 557
 Db 538 vsvdreespkvl-ita 553
 RESULT 10
 AAB43038
 ID AAB43038 standard; Protein; 560 AA.
 XX AAB43038;
 XX 08-FEB-2001 (first entry)
 XX Human ORFX ORF2802 polypeptide sequence SEQ ID NO:5604.
 XX Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
 KW vulnery; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
 KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
 KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
 KW thrombosis; contraceptive.
 XX Homo sapiens.
 XX WO200058473-A2.
 XX 05-OCT-2000.
 XX 31-MAR-2000; 2000WO-US08621.
 XX 31-MAR-1999; 99US-0127607.
 PR 02-APR-1999; 99US-0127636.
 PR 05-APR-1999; 99US-0127728.
 PR 30-MAR-2000; 2000US-0540763.
 XX (CURA-) CURAGEN CORP.
 XX Shinkets RA, Leach M;
 PI WPI; 2000-602362/57.
 DR N-PSDB; AAC77247.
 XX Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease -
 XX Claim 11; Page 4791-4792; 5507pp; English.
 CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytostatic; hepatotropic; vulnery;
 CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
 CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
 CC antithyroid; and antianemic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating

Db 458 ptfvnmvcsedldgdltpfivfrlvrevwqalpdlfavlgllaagvtlllpctk 517
 QY 514 GTPLPDTI 521
 Db 518 gvalpetm 525
 RESULT 12
 AAB49401
 ID AAB49401 standard; Protein; 540 AA.
 AC AAB49401;
 XX
 DT 02-MAR-2001 (first entry)
 DE Murine organic anion transporter 6.
 XX Murine; organic anion transporter 6; mOATP6; cancer; inflammation;
 KW cardiovascular disease; central nervous system disorder; kidney disease;
 KW liver disease; autoimmune disease.
 XX Mus sp.
 OS
 XX WO200070048-A1.
 PN 23-NOV-2000.
 PD
 PF 15-MAY-2000; 2000WO-US13316.
 PR 14-MAY-1999; 99US-0134137.
 PR 12-MAY-2000; 2000US-0570233.
 XX (SMIK) SMITHKLINE BEECHAM CORP.
 PA (SMIK) SMITHKLINE BEECHAM PLC.
 PA
 PI Feild J, Yue L, Ellens H;
 XX
 DR WPI; 2001-016235/02.
 DR N-PSDB; AAC83979.
 XX Murine organic anion transporter 6 polypeptide, useful for identifying
 PT agonists/antagonists that are useful in treatment of cancer, kidney
 PT disease, autoimmune disease, inflammation and cardiovascular disease
 XX
 PS Claim 2; Page 28-29; 32pp; English.
 XX The present sequence is murine organic anion transporter 6 (mOATP6).
 CC mOATP6 protein is useful for screening compounds which inhibit or
 CC stimulate the function of mOATP6 and also compounds that neither agonise
 CC nor antagonise OATP6. The identified agonists and antagonists are useful
 CC for prevention and treatment of human diseases, including cancer,
 CC inflammation, cardiovascular disease, central nervous system disorders,
 CC kidney diseases, liver disease and autoimmune diseases.
 XX
 SQ Sequence 540 AA;
 Query Match 25.1%; Score 724.5; DB 22; Length 540;
 Best Local Similarity 34.4%; Pred. NO. 4e-70;
 Matches 187; Conservative 89; Mismatches 237; Indels 31; Gaps 13;
 QY 4 YDEVTAFLEWGPQ--RLIFFELLSAIIIPNGFTGLSSVVELIATPEHRCRPDA-ANLS- 59
 Db 3 feellhkvvgfqlrnlvialprflpmhf--llplfmaavpahcalpdpapanlsh 60
 QY 60 -SAWRNHTVPLRUDGREVPHSCRRYRLATIANFSAIGLEPGRDVDLGLQLEQ----SCL 114
 Db 61 qdlwlkthip-retdsf--ssclrfaypqalpnlvtlgtrev--ynsgepegepltpcs 114
 QY 115 DGWEFSQDYLSTIVTEWNLVCEDDWKAPLITISLFFVGLGSLFSGQLSDFGRKNVLF 174
 Db 115 qgweydrsfssfiataedwldvcqgrlgnkvstctcfignllgavvyeylsdrfgrrill 174

QY 175 VTMQMOTGFSFLOIFSKNFEMFVFLVFLVGMGOISNVAFAFVLGTTEILGKSVRIIFSLG 234
 Db 175 vayvstlalglmsaasvnyimfvtrmltgsalagftiivlpolewidvehrtvagvis 234
 QY 235 VCIFYAFGYMVLPLFAYFIRDMRLMLVALTMPGVLCAVMWFIPSPRWLISQGRFEAE 294
 Db 235 t-ftwtggvllltlvgyllrswrwilllaatlpcvpgiliswwpesarwlltqgrveeak 293
 QY 295 VIIRKAAKANGIVVPSTIPDPSELDLSSKKQ--QSHNTLDLLRTWNIRMTIMSILMM 352
 Db 294 kylisicaklngrpisedsisealnkvitmervsgrpsyldlfrtsqirhvsicommwuf 353
 QY 353 TISVGYEGLSLDTPNLHGDIIVNCFLSAMVEVPAYYLAWLMLLQYLPRRYS-----MATAL 407
 Db 354 gvnfsyyglltdasgigtvyqtqllfgavevpskitvfflvrlgrlriteagmlatal 413
 QY 408 FLGGSVLLFMQLVPPDLYLATVLMVVGKFGVTAAFSMWVYVTAELPYTVVRNMGVYSS 467
 Db 414 tfgisl-----lvssdtkswitalvvgkafseaaftaylftselypvlrqtgmfta 468
 QY 468 TASFGLSTLSPYFVYVLCAYDRELPYILMGLSLTILTAITLFLPESFGTLPDPTIDQMLRV 527
 Db 469 ligrlgaslaplvllvgvlllpkiayggisflaactvlllpetkkaqlpetiqdver- 527
 QY 528 KGKM 531
 Db 528 kgk 531
 RESULT 13
 AAY51249
 ID AAY51249 standard; Protein; 535 AA.
 XX AAY51249;
 AC AAY51249;
 DT 14-APR-2000 (first entry)
 DE Rat liver anion transporter protein OAT2.
 XX OAT2; anion transporter; liver; rat; dicarboxylic acid; prostaglandin;
 KW non-steroid anti-inflammatory agent; anti-tumor.
 XX Rattus sp.
 OS
 PN JP11346779-A.
 XX 21-DEC-1999.
 XX 03-JUN-1998; 98JP-0169174.
 XX 03-JUN-1998; 98JP-0169174.
 XX (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.
 DR WPI; 2000-109694/10.
 DR N-PSDB; AAZ44679.
 XX A liver-specific org, anion transporter - can transport various
 PT substances including dicarboxylic acids
 XX Claim 2; Page 5-6; 13pp; Japanese.
 XX This invention describes a novel rat liver-specific org. anion
 CC transporter OAT2. The liver-specific org. anion transporter OAT2
 CC can transport various substances including dicarboxylic acids,
 CC prostaglandins, non-steroid anti-inflammatory agents and anti-tumor
 CC agents. This sequence represents the rat liver anion transporter OAT2
 CC protein described in the method of the invention.
 XX Sequence 535 AA;
 SQ
 Query Match 24.9%; Score 719; DB 21; Length 535;

```
Best Local Similarity 34.18; Pred. No. 1.6e-69;
Matches 183; Conservative 96; Mismatches 232; Indels 26; Gaps 14;

QY 4 YDEVTAFLGEGGFFO--RLIFLLSASTIPNGFTGLSSVFLIATPEHRCRVPDA--ANLS- 59
Db 3 fecilidkvggfgpqlrlvlnmalprmlphf--llpvmavpahcalpgapanish 60
QY 60 -SARNHTVPLRLDRGVRPHSCHRRYL--ATIANFSALGLEPGRDVLDGLOEQE---SC 113
Db 61 qdlwleahlp-retgdsf--sscirfayptqvpntv-lgtevans---gegepeptvpc 113
QY 114 LDGWEFSQDVLSTIVTWNLCVDEDDKAPLTISLFPVGVLLGSFISQISDRFGKRNVL 173
Db 114 sgdweydrsefsstatedwldvcgqrlnkstctffiglvgavvyglsdrfgrll 173
QY 174 FVTMGWQGFSLQIFSKNFEMFVFLVGVGMQISNVAAFLVGTILCKSVRIIFSTL 233
Db 174 lvayvssilvlgmsaasinyimfvvtrltgsalagftiivplelewdvehrtvagvi 233
QY 234 GVCIFYATGVVLPFLPAYFIRDRMLLVALTMPGVLCVALWFIPESPRWLISQGRPEEA 293
Db 234 st-vfwsdgvgvillalvgylirswrlllaatlpcvpqgiswvwpesarwlltqgrveea 292
QY 294 EVIIRKAANKANGIVVPSTIFPDSQLSSKKQ--QSHNILDLLRTWNIRMTWIMSLW 351
Db 293 kylliscaklngprvgeslsqealnnavvtneralqpsyldfrftrtsqlrhslccmmvw 352
QY 352 MTISVGVYGLSDTNPNGHDFVNCFTSAMVEVPAYVLAMLLQYLPDRYSMATALFLGG 411
Db 353 fgvnfsyvgltldvsglgnvyqqlfgavelpsklmvylvrlgrlgriteagml--g 410
QY 412 SVLLF--MQLVPPDLYLATVLMVGMKFGVTAAPFSMVVVTAEYPTVVRNMGVGSSTA 469
Db 411 aaltgtsllvslctskwtalvvvgkafseaaftaylftselypvtvltqtlgitalm 470
QY 470 SRLGSILSPFYVYGAYDRFLPYILMGLSLTLTALFLFLPESFGTPLPTIDOMLR 526
Db 471 grigaslaplaalldgwwlllpkayvgialvaactallipetkkaqlbetiqdver 527

RESULT 14
AAR77676
ID AAR77676 standard; protein; 556 AA.
AC AAR77676;
DT 10-SEP-1996 (first entry)
DE Rat OCT-1 protein.
KW Rat; OCT-1; transporter protein; cationic; xenobiotic; pharmaceutical;
KW blood; liver; kidney; epithelial cell; intestine; tetraethylammonium;
KW proximal renal tubule cell; intestine; enterocyte; transgenic; renal;
KW biliary; excretion; resorption; modulator; uptake.
OS Rattus rattus.
XX
XX
XX Key Location/Qualifiers
XX Domain 20..46
XX Modified-site /note= "transmembrane domain"
XX Modified-site 71
XX Modified-site 97
XX Modified-site /note= "putative N-glycosylation site"
XX Modified-site 113
XX Modified-site /note= "putative N-glycosylation site"
XX Domain /note= "putative N-glycosylation site"
XX Domain 154..171
XX Domain /note= "transmembrane domain"
XX Domain 178..197
XX Domain /note= "transmembrane domain"
XX Domain 243..260
XX Domain /note= "transmembrane domain"
XX Domain 267..283
```

```
FT Domain /note= "transmembrane domain"
FT 350..366
FT /note= "transmembrane domain"
FT 380..398
FT /note= "transmembrane domain"
FT 406..425
FT /note= "transmembrane domain"
FT 432
FT Modified-site
FT 435..452
FT /note= "transmembrane domain"
FT 463..485
FT /note= "transmembrane domain"
FT 494..514
FT /note= "transmembrane domain"
XX DE4424577-AL.
XX 18-JAN-1996.
XX 13-JUL-1994; 94DE-4424577.
XX 13-JUL-1994; 94DE-4424577.
XX (FARH ) HOECHST AG.
XX Gorboulev V, Gruendeman D, Koepsell H;
XX WPI; 1996-069422/08.
XX N-PSDB; AAT08702.
XX Transporter protein for cationic xenobiotic(s) and pharmaceuticals,
XX and related DNA and transformed cells - used e.g. to assess
XX excretion and resorption of cationic cpds.
XX Claim 1; Fig 2A1; 13pp; German.
XX This is the amino acid sequence of the rat OCT-1 protein, a new
XX transporter protein able to transport cationic xenobiotics and
XX pharmaceuticals from the blood into liver or kidney epithelial
XX cells or from the intestine. The gene was isolated by infecting
XX a rat kidney gene library into Xenopus laevis oocyte and isolating
XX clones conferring uptake of 14C-tetraethylammonium. One clone
XX designated OCT-1 was isolated. Expression of the gene was detect
XX in proximal renal tubule cells, in liver epithelial cell and in
XX intestinal enterocytes. The DNA can be used to generate transgenic
XX cells for use in vitro test for renal/biliary excretion or
XX intestinal resorption of xenobiotics and pharmaceuticals. The protein
XX or cells expressing it can also be used to isolate modulators that
XX block uptake of pharmaceutical by the renal tubules.
XX Sequence 556 AA;
XX
XX Query Match 24.7%; Score 712; DB 17; Length 556;
XX Best Local Similarity 33.8%; Pred. No. 9.7e-69;
XX Matches 193; Conservative 96; Mismatches 230; Indels 52; Gaps 14;
QY 1 MRDYDEVTAFLGEGGFFOQLIFLL---SASITPNQFTGLSSVFLIATPEHRCRVPDAAN 57
Db 1 mptvddvlegvgfegfkgqafllcllclisaslapi-yvgi--vflgftpgqcpvgvae 57
QY 58 LSS--AWR-----NHTVP-LRLDRGVRPHSCHRRYL-----ATIANFSAL 95
Db 58 lsqrcgwsqaeeelnvtpvgigpsdeasflsqcmryevdwngstldcvplssivansql 117
QY 96 GLEPGRDVLDGLOEQESCLDGWEFSQDVLSTIVTWNLCVDEDDKAPLTISLFFGVLL 155
Db 118 plpg-----cehgwy--dtpgssivtefnlvcgdawkvdlfqscvnlgfll 162
QY 156 GSFISGQLSDRFRKKNVLFVTMGWQISNVAAFLVGTILCKSVRIIFSTL 215
Db 163 gslvvgyladrfrgrklcllvtlvtvsvglavapdytsmlilfrllqgmvsqswsyy 222
```

QY 216 VLGTIELGKSVRIIFSTGLGYCIFAF--GYMVLPLFAYFIRDRMLLVALTMGVLVAL 273
 Db 223 tlltefvggyr---rttallyqmaftvglgvayalpdrwlglaavslptflily 279
 QY 274 WFTIPESPRWLISQGRPEAEVIRKAANGIVVPSTIFDSELDLSSKKQOSHIL 333
 Db 280 ywfesprwllsqktrtravimeqladkngkvpdpadlmlcleedasekr--spsfad 337
 QY 334 LLRTWNRMTVIMSIMLWMTISVGYFGLSLDTPNLHGDFVNCFLSAMVEVPAYVLA 393
 Db 338 lfrtptlrkhtvilmlywfscavlygglmhvgatganlylaffyslvefpaafilvt 397
 QY 394 LQYLPRYSMATALFLGGVLLFMQLVPPDLXYLAVLVLMVWKGFGVTAFAFSWYVY 453
 Db 398 idrigrypiaasnlvtgaacclimifiphelhwnltacigrmgatvlgmclvnael 457
 QY 454 YPTVVRNMGVGVSTASRLGSLSPYFY--LGAYDRFLPVILMGSTILTALTILFLPES 512
 Db 458 yptfirnlgmvmvscaldlgiftpmvfrlmevwqalpiliifvgigtagamtililpet 517
 QY 513 FGTPIDPTIDQMLRVKGMKHKTPSHTRMLK 543
 Db 518 kgvalpetieaeal-grrksakentiyiq 547

RESULT 15
 AAY92902
 ID AAY92902 standard; Protein; 542 AA.
 AC AAY92902;
 XX
 XX
 XX 26-SEP-2000 (first entry)
 DE Human cerebral organic anion transporter OAT3 protein.
 KW Human; organic anion transporter protein; OAT3; cerebral tissue.
 OS Homo sapiens.
 XX
 XX WO200017237-A1.
 PN 30-MAR-2000.
 PD
 XX 20-SEP-1999; 99WO-JP05120.
 XX 18-SEP-1998; 98JP-0265126.
 XX (NISC-) JAPAN SCI & TECHNOLOGY CORP.
 XX Endou H, Sekine T, Kusuvara H;
 XX WPI; 2000-283546/24.
 DR N-PSDB; AAA11146.
 XX Organic anion transporter protein OAT3 expressed in cerebral tissue,
 PT antibodies to it and gene encoding it for study of cerebral transport
 PT of anions including drug molecules
 XX
 XX Claim 2; Page 18-20; 48pp; Japanese.
 XX
 CC This sequence represents a human organic anion transporter protein OAT3.
 CC The sequence is used to regulate the transport of anions in cerebral
 CC tissue, including drug molecules, and regulation of this transport.
 XX
 SQ Sequence 542 AA;

Query Match 24.0%; Score 693; DB 21; Length 542;
 Best Local Similarity 32.9%; Pred. No. 1.1e-66;
 Matches 179; Conservative 92; Mismatches 229; Indels 44; Gaps 13;
 QY 4 YDEVTAFLGHWGPFQRLIFLLSASIPNGFTGLSSFLIATPEHRCRVPDAANLSSAWR 63

Db 3 fseildrvsmghbfqlhvailglpilmanhllqiftaatpvhrcrphnas-tgpw- 60
 QY 64 NHTVPLRLDRGREVPHSCRRY---RLATIANFALGLPEGRDVLGQLEQESCLDGHEFS 120
 Db 61 --vlpmg-pngk--percirfwhppnaslpndtqramep-----cidgw--- 99
 QY 121 QDYVLSL---IVTEWNLVCEDDWKAPLTISLFFVGVLLGSFISGQLSDFRFGKKNVLFVTM 177
 Db 100 --vynstkdsivtewdlvcnshklkemaqsfmagilvgldslrfrpiltcsy 157
 QY 178 GMOTGSFSLQIFSKNFEFVFLVFLVGMGOISNYVAAVFLGTEILGKSVRIIFST-LGVC 236
 Db 158 lllaaagsgaafspftiyvmvfrlclgfgisgitlsvilnvwvptrmraimstalgyc 217
 QY 237 IFYAFGVWVLPFAYFIRDRMLLVALTMGVLVALWVFIPESPRWLISQGRPEAEVI 296
 Db 218 --ytfqgilplglayalpqrwrlqtlvsipffvflsswwtpesirwlvsgksealki 275
 QY 297 IRKAAKANGIVVPSTIFDSELDLSSKKQOS-----HNILDLRLTWNIRMTVIMSILW 351
 Db 276 lrrvavfngkkeeerlsleek-lnlqkeislakakytasdlfrimprrmtfclslaw 334
 QY 352 MTISVGYFGLSLDTPNLHGDFVNCFLSAMVEVPAYVLAWLLLOYLPRYSMATALFLGG 411
 Db 335 fatgfayyslamgveefgvnlyilqilfggvdpakfifilsyigrhttgaaallag 394
 QY 412 SVLLFMQLVPPDLXYLATVLVWVGKFGVTAFAFSWYVYTAELYPVVRNMGVGVSTASR 471
 Db 395 gailaltfvpldlqtrvtvrlavfvgkclsssfcliytselyptvirqtgmvgvsnlwr 454
 QY 472 LGSILSPVYVYLGAYDRFLPVILMGSLTILTALTILFLPESFGTPLDPTIDQM---LRV 527
 Db 455 vgsmsvplvkitgevqfipniygitallggsaalflpetlloqplpetiedlenwslra 514
 QY 528 KGMK 531
 Db 515 kkp 518

Search completed: August 16, 2001, 13:58:26
 Job time: 208 sec

.....

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 16, 2001, 13:58:52 ; Search time 20.69 Seconds
(without alignments)
554.317 Million cell updates/sec

Title: US-09-521-195-3
Perfect score: 2883
Sequence: 1 MRDYDVTAFLGEGWGFQRL.....HTRMLKDGERTILKSTAF 557

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pap.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pap.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pap.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pap.*
5: /cgn2_6/ptodata/2/iaa/PCITUS_COMB.pap.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	726	25.2	555	3	US-08-501-572-3
2	726	25.2	555	3	US-09-040-444-3
3	718	24.9	556	3	US-08-501-572-1
4	718	24.9	556	3	US-09-040-444-1
5	714.5	24.8	553	3	US-08-501-572-2
6	714.5	24.8	553	3	US-09-040-444-2
7	652.5	22.6	537	2	US-08-647-397-2
8	234	8.1	494	2	US-09-031-392-5
9	234	8.1	494	4	US-09-299-549-5
10	208	7.2	492	2	US-08-355-844-3
11	208	7.2	492	5	PCT-US95-16126-3
12	203	7.0	493	2	US-09-031-392-10
13	203	7.0	493	4	US-09-299-549-10
14	198	6.9	109	2	US-08-647-397-4
15	193	6.7	509	2	US-09-031-392-6
16	193	6.7	509	4	US-09-299-549-6
17	186	6.5	500	2	US-09-031-392-7
18	186	6.5	500	4	US-09-299-549-7
19	184	6.4	563	2	US-09-031-392-2
20	184	6.4	563	4	US-09-299-549-2
21	175	6.1	488	2	US-08-928-692-11
22	168.5	5.8	534	2	US-09-031-392-4
23	168.5	5.8	534	4	US-09-299-549-4
24	155	5.4	383	2	US-09-031-392-3
25	155	5.4	383	4	US-09-299-549-3
26	153.5	5.3	524	2	US-08-928-692-12
27	149	5.2	584	2	US-08-928-692-13

28	131	4.5	488	2	US-08-928-692-10	Sequence 10, Appl
29	110	3.8	473	1	US-08-597-236-13	Sequence 13, Appl
30	110	3.8	473	1	US-08-746-682A-13	Sequence 13, Appl
31	104.5	3.6	1835	3	US-08-836-325-15	Sequence 15, Appl
32	101.5	3.5	341	1	US-08-423-564-5	Sequence 5, Appl
33	97	3.4	342	3	US-08-785-928-1	Sequence 1, Appl
34	97	3.4	342	3	US-08-728-603-17	Sequence 17, Appl
35	97	3.4	916	2	US-08-928-692-58	Sequence 58, Appl
36	94.5	3.3	390	5	PCT-US92-02091-8	Sequence 8, Appl
37	93.5	3.2	467	2	US-08-805-118-3	Sequence 3, Appl
38	93.5	3.2	480	2	US-08-724-394A-9	Sequence 9, Appl
39	93	3.2	560	2	US-08-805-118-4	Sequence 4, Appl
40	92.5	3.2	419	1	US-08-385-186-2	Sequence 2, Appl
41	92.5	3.2	568	4	US-08-637-823B-27	Sequence 27, Appl
42	92.5	3.2	584	4	US-08-637-823B-28	Sequence 28, Appl
43	92	3.2	429	2	US-08-677-049-5	Sequence 5, Appl
44	92	3.2	503	4	US-09-068-195-24	Sequence 24, Appl
45	92	3.2	548	3	US-08-903-139B-28	Sequence 28, Appl

ALIGNMENTS

RESULT 1
US-08-501-572-3
; Sequence 3, Application US/08501572
; Patent No. 6063623

; GENERAL INFORMATION:
; APPLICANT: Koepsell, Hermann
; APPLICANT: Grundeman, Dirk
; APPLICANT: Gorboulev, Valentin
; TITLE OF INVENTION: Transport protein Which Effects The
; TITLE OF INVENTION: Transport Of Cationic Xenobiotics and/or Pharmaceuticals,
; TITLE OF INVENTION: DNA Sequences Encoding It And Their Use.
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan,Henderson,Farabow,Garrett & Dunner
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin: Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/501,572
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Toohey, Kimberlin M
; REGISTRATION NUMBER: 35,391
; REFERENCE/DOCKET NUMBER: 02481.1453-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)408-4000
; TELEFAX: (202)408-4400
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 555 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-501-572-3

Query Match 25.2%; Score 726; DB 3; Length 555;
Best Local Similarity 35.0%; Pred. No. 5.2e-68;
Matches 187; Conservative 98; Mismatches 223; Indels 26; Gaps 14;
QY 5 DEVTAFLGEGWGFQRLIFF---LISASITPNGFTGLSSVFLIATPEHRCRVPDANLUS-- 59

```

; LENGTH: 555 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-040-444-3

Query Match      25.2%; Score 726; DB 3; Length 555;
Best Local Similarity 35.0%; Pred. No. 5.2e-60;
Matches 187; Conservative 98; Mismatches 223; Indels 26; Gaps 14;

QY 5 DEVTAFLGEGWPFQRLIFF---LLSASTIPNGFTGLSSVFLIATPEHRCRVPDAANLS-- 59
DB 6 DDVLEHGGEFFHFKQMEFFLALLSATFAP1-YVGI--VFLGTPDHRCSRPGVAELSRL 62
QY 60 SAWR-----NHTVPLRLDGRVPHSCRRYRLA-TIANESALGLEPGRDVDLQLEQESC 113
DB 63 COWSPAELNVTVPGPAGEASPRQCRRYEVDWQNSTFDCVDPLASLDTNRSRLPLGPC 122
QY 114 LDGWEFSQDVLSTIVTWNLVCEDDWKAPLTISLFFVGVLLGSFISGQSLDRGRKNVL 173
DB 123 RDGWY--ETPGSSIVTEFNLCANSMWLDLQSSVNVGFFIGSMISGYIADRFGRKLCL 180
QY 174 FVTGMQTFGSFLQIFSKNFEMFVFLVLMGMOISNYVAFAVLGTGLGKSVRIIEFTL 233
DB 181 LTTVLINAAAGVLAISPTYTWMFLIRLOGLVSKAGWLGIIILITEFVGGRYR---RTV 237
QY 234 GVC--IFYAFGYMVLPLPAYFIRDMRMLLVALTMPGVLCVAMWFFIPESRWLISQGRFE 291
DB 238 GIFQVAYTVGLLVAGVAYALPHWRWLQFTVALPNEFFLLYWCIPESRWLISQNKNA 297
QY 292 EAEVIIRKAAKANGIVVPSTIFDPSELQDLSKKQSHNLDLLRTWNIRMTIMSLW 351
DB 298 EAMRIIKHIAKKNGKSLPASL-QRLLEEETGKK-LNPSFLDLVTRTPQIRKHTMILMYN 355
QY 352 MTISVGVFGLSLDTPNLHGD-IFVNCFLSAMVEVPAYVLAWLLIQLYPRRYSMATLFLG 410
DB 356 FTSSVLYOGLIMHM-GLAGDNIYLDFFYSALVEPPAAFMIIILIDRIGRRYPWAAASNMYA 414
QY 411 GSVLLFMQLPDPDLYLYLATVLMVKGKGTAAASWVYVYTAELYPTVVRNMGVGVSTAS 470
DB 415 GAACLASVFIPGDLQWLKIIISCLGRMGITMAYEIVCLVNAELYPTFIRNLGVHCSSMC 474
QY 471 RLGSILSPFYVY-IGAYDRFLPYLMGSLTILTAITLFLPESFGPLPDTIDQ 523
DB 475 DIGGIITPFLVYRLTNWLELPLMWFGVLVAGGLALLLPETKALPETIEE 528

```

RESULT 2

```

US-09-040-444-3
; Sequence 3, Application US/09040444
; Patent No. 6063766
; GENERAL INFORMATION:
; APPLICANT: Koepsell, Hermann
; APPLICANT: Grundeman, Dirk
; APPLICANT: Gorboulev, Valentin
; TITLE OF INVENTION: Transport protein which effects the
; TITLE OF INVENTION: Transport of cationic xenobiotics and/or pharmaceuticals,
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner, L.L.P.
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/040.444
; FILING DATE: March 18, 1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: O'Connor, Steven P
; REGISTRATION NUMBER: 41,225
; REFERENCE/DOCKET NUMBER: 2481.1453-01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)408-4000
; TELEFAX: (202)408-4400
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:

```

RESULT 3

```

US-08-501-572-1
; Sequence 1, Application US/08501572
; Patent No. 6063623
; GENERAL INFORMATION:
; APPLICANT: Koepsell, Hermann
; APPLICANT: Grundeman, Dirk
; APPLICANT: Gorboulev, Valentin
; TITLE OF INVENTION: Transport protein which effects the
; TITLE OF INVENTION: Transport of cationic xenobiotics and/or pharmaceuticals,
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30

```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/501,572
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Toohy, Kimberlin M
; REGISTRATION NUMBER: 35,391
; REFERENCE/DOCKET NUMBER: 02481.1453-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)408-4000
; TELEFAX: (202)408-4400
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 556 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-501-572-1

```

```

Query Match      24.9%; Score 718; DB 3; Length 556;
Best Local Similarity 34.0%; Pred. No. 3.6e-67;
Matches 194; Conservative 96; Mismatches 229; Indels 52; Gaps 14;

QY 1 MRDYDEVTAFLGEMGPFQRLIFELL--SASIIPIGNGFTGLSSVFLIATPEHRCRVPDAAN 57
Db 1 MPTVDVLEQVGEFGWQKQAFLLCLISASLAPI-YVGI--VFLGFTPGHYCONPGVAE 57

QY 58 LSS--AWR-----NHTVP-LRLRDGREVPHSCRRYL-----ATIANFSAL 95
Db 58 LSQRCGWSQAEEELNTPVGLGPSDEASFLSQCMRYEVDWNGQSTLDCVDP LSSLVANRSQL 117

QY 96 GLEPGRDVLDGLEQSCLDGWFSDVYLSITVTEWNLVCEDDWKAPLISLFFVGVLL 155
Db 118 PLGP-----CEHGWY--DTPGSSIVTEFNLCVGDWAKVDFLQSCVNLGFFL 162

QY 156 GSPISGQLSDRFGKKNLVFTMGMTGTFQIFSKNPFVFLVLCMGQISNYYAAF 215
Db 163 GSVLVGYIADRGKRLCLLVTLVTSVGLVAVAPDYSMTLLFRLLQGMVSKGWSVSGY 222

QY 216 VLGTILGKSVRIIFSTLGVCFYAF--GYMVLPLFAFYFIRDMRLMLVALTMPGVLCVAL 273
Db 223 TLITEFVSGSYR---RTTALYQMAFTVGLVGLAGVAYAPDWRWLQAVSLPTFLFLY 279

QY 274 WFFIPESRWLISQGRFEAEVIRKAAKANGIVVPSTIFDPSLQDLSSKKQOQSHNILD 333
Db 280 YWFPESRWLISQGRFTRAVRIMEQIAQKNGKVPADLKLMLCLEEDASEKR--SPSEAD 337

QY 334 LLRTNIRMTVIMSIMLWMTISVGYFGLSLDTPNLHGDIFVNCFLSAMVEVPAYVLAWL 393
Db 338 LFTPNLRKHTVILMTLWLFSCAVLYQGLIMHVGTAGANLYLDFYSSLVFEPPAAFIILVT 397

QY 394 LQYLPTRYSMATLFLGGSVLLFMQLVPPDLYLATLVLMVWGKFGVTAAFSMVYVTAEL 453
Db 398 IDRIGRIYPIAASNLVTGAACLLMIFIPHELHNLVTLACLGRMGATIVLEMVCLVNAEL 457

QY 454 YPTVIRNMGVGSSTASRLSILSPFYV-LGAYDRFLPYILMGSITILTALTFLFLES 512
Db 458 YPTFIRNLGMVCSALCDLGGITFPFVPRMLMEVMOALPLILFGLVGLTAGANTLLLPET 517

QY 513 FGTPLPDITDMLRVKGMKHKRTPSPHRLMK 543
Db 518 KGVALPETIEEENL--GRRKSKAKENTIIYLQ 547

```

```

RESULT 4
US-09-040-444-1
; Sequence 1, Application US/09040444
; Patent No. 6063766
; GENERAL INFORMATION:
; APPLICANT: Koepsell, Hermann
; APPLICANT: Grundeman, Dirk

```

```

; APPLICANT: Gorboulev, Valentin
; TITLE OF INVENTION: Transport protein Which Effects The
; TITLE OF INVENTION: Transport of Cationic Xenobiotics and/or Pharmaceuticals,
; TITLE OF INVENTION: DNA Sequences Encoding It And Their Use.
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan,Henderson, Farabow, Garrett & Dunner, L.L.P.
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/040,444
; FILING DATE: March 18, 1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: O'Connor, Steven P
; REGISTRATION NUMBER: 41,225
; REFERENCE/DOCKET NUMBER: 2481.1453-01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)408-4000
; TELEFAX: (202)408-4400
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 556 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-040-444-1

```

```

Query Match      24.9%; Score 718; DB 3; Length 556;
Best Local Similarity 34.0%; Pred. No. 3.6e-67;
Matches 194; Conservative 96; Mismatches 229; Indels 52; Gaps 14;

QY 1 MRDYDEVTAFLGEMGPFQRLIFELL--SASIIPIGNGFTGLSSVFLIATPEHRCRVPDAAN 57
Db 1 MPTVDVLEQVGEFGWQKQAFLLCLISASLAPI-YVGI--VFLGFTPGHYCONPGVAE 57

QY 58 LSS--AWR-----NHTVP-LRLRDGREVPHSCRRYL-----ATIANFSAL 95
Db 58 LSQRCGWSQAEEELNTPVGLGPSDEASFLSQCMRYEVDWNGQSTLDCVDP LSSLVANRSQL 117

QY 96 GLEPGRDVLDGLEQSCLDGWFSDVYLSITVTEWNLVCEDDWKAPLISLFFVGVLL 155
Db 118 PLGP-----CEHGWY--DTPGSSIVTEFNLCVGDWAKVDFLQSCVNLGFFL 162

QY 156 GSPISGQLSDRFGKKNLVFTMGMTGTFQIFSKNPFVFLVLCMGQISNYYAAF 215
Db 163 GSVLVGYIADRGKRLCLLVTLVTSVGLVAVAPDYSMTLLFRLLQGMVSKGWSVSGY 222

QY 216 VLGTILGKSVRIIFSTLGVCFYAF--GYMVLPLFAFYFIRDMRLMLVALTMPGVLCVAL 273
Db 223 TLITEFVSGSYR---RTTALYQMAFTVGLVGLAGVAYAPDWRWLQAVSLPTFLFLY 279

QY 274 WFFIPESRWLISQGRFEAEVIRKAAKANGIVVPSTIFDPSLQDLSSKKQOQSHNILD 333
Db 280 YWFPESRWLISQGRFTRAVRIMEQIAQKNGKVPADLKLMLCLEEDASEKR--SPSEAD 337

QY 334 LLRTNIRMTVIMSIMLWMTISVGYFGLSLDTPNLHGDIFVNCFLSAMVEVPAYVLAWL 393
Db 338 LFTPNLRKHTVILMTLWLFSCAVLYQGLIMHVGTAGANLYLDFYSSLVFEPPAAFIILVT 397

QY 394 LQYLPTRYSMATLFLGGSVLLFMQLVPPDLYLATLVLMVWGKFGVTAAFSMVYVTAEL 453
Db 398 IDRIGRIYPIAASNLVTGAACLLMIFIPHELHNLVTLACLGRMGATIVLEMVCLVNAEL 457

QY 454 YPTVIRNMGVGSSTASRLSILSPFYV-LGAYDRFLPYILMGSITILTALTFLFLES 512
Db 458 YPTFIRNLGMVCSALCDLGGITFPFVPRMLMEVMOALPLILFGLVGLTAGANTLLLPET 517

QY 513 FGTPLPDITDMLRVKGMKHKRTPSPHRLMK 543
Db 518 KGVALPETIEEENL--GRRKSKAKENTIIYLQ 547

```

QY 454 YPTVNRNMGVGVSTASRLGSLSPYFY-LGAYDRFLPYILMGSLTILTLFLPES 512
 Db. 458 YPTFIRNLNMGVCSALCDLGGITFPFVRLMEVWQALPLILFVGLGTAGAMTILLPET 517
 QY 513 FGTPLPDTIDQMLRVKGMKHKRTPSHRMLK 543
 Db 518 KGVALPETIEEAEHL-GRKSKAKENTYLO 547

RESULT 5
 US-08-501-572-2
 ; Sequence 2, Application US/08501572
 ; Patent No. 6063623
 ; GENERAL INFORMATION:
 ; APPLICANT: Koepsell, Hermann
 ; APPLICANT: Grundeman, Dirk
 ; APPLICANT: Gorboulev, Valentin
 ; TITLE OF INVENTION: Transport protein which effects the
 ; TITLE OF INVENTION: Transport of cationic xenobiotics and/or pharmaceuticals,
 ; TITLE OF INVENTION: DNA sequences encoding it and their use.
 ; NUMBER OF SEQUENCES: 6
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner
 ; STREET: 1300 I Street, N.W., Suite 700
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: USA
 ; ZIP: 20005-3315
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent in Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/501,572
 ; FILING DATE:
 ; CLASSIFICATION: 424
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Toohey, Kimberlin M
 ; REGISTRATION NUMBER: 35,391
 ; REFERENCE/DOCKET NUMBER: 02481.1453-00000
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (202)408-4000
 ; TELEFAX: (202)408-4400
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 553 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: peptide
 ; US-08-501-572-2

Query Match 24.8%; Score 714.5; DB 3; Length 553;
 Best Local Similarity 35.4%; Pred. No. 8.4e-67;
 Matches 194; Conservative 87; Mismatches 216; Indels 51; Gaps 13;

QY 1 MRDYDEVTAFLCEWGFOR---LIFLLSASLIPNGFTGLSSVFLIATPEHRCRVPDAAN 57
 Db 1 MPTVDLLEQVGESEWFOKQAFILICLLSAFAP---ICVGIIVFLGFTPDHHCQSPGVAE 57
 QY 58 LSS--AWR-----NHTVPLRLRDGVEPHSCRRYR-----LATIA-NFSALG 96
 Db 58 LSQRGWSPEBELNTPVGLPAGEAFLGQCRRYEVDWQNSALSCVDPPLASLATNRSLP 117
 QY 97 LEPRGVDLGLQEQESCLDGHEFSDDVYLSIVTEWNLVCEDDWKAPLTISLFFVGVLLG 156
 Db 118 LGP-----CDQGWY---DTGSSIVTEFNLCADSKWLDLFSCLNAGFFFG 162
 QY 157 SPISQLSDRFKKNVFLVTMGMTQTFGLQIFSKNFEMFVFLVLMGMGQISNVAAVF 216

Db 163 SLGVGVFADRFGRKLCLLGTVLVNAVSGVLMAFSPNYSMLLFRLLQGLVSKGNWAGYT 222
 QY 217 LGTEILGKSVRIIFSTLGVCFYAF--GYMVLPLFAFYFIRDMRLMLVALTMPGVLCAVLM 274
 Db 223 LITEFVGSGSR---RTVAIMYQMAFTVGLVALTGLAYALPHWRLQLAVSLPTFFELLY 279
 QY 275 WFIPESPRWLISQGRFEAEVIRKAANGIVVPTIFDPSLQDLSSKKQOQSHNILD 334
 Db 280 WCVPESPRWLLSQKRNTAEIKIMDHIAQKNGKLPADLKLMSLEEDVTEK--LSPSFADL 337
 QY 335 LRTWNIRMTVIMSIMLWMTISVGYFGLSLDTPNLHGDI FVNCFLSAMVEVPAYVLAWL 394
 Db 338 FRTPRLRKRTFILMYLWFTDSVLYQGLIHMGATSNLYDLFDLYSALVEIPGAFILITI 397
 QY 395 QYLPRRYSMATALFLGSGVLLFMQLVPPDLYLATVLMVVGKGVTAAFSMVYVYAE 454
 Db 398 DRVGRIPMAVSNLLAGAACLVI-FISPDHLHLNIIMCVGRMGITIAIOMICLVNAELY 456
 QY 455 PTVVRNMGVSVSTASRLGSLSPYFY-LGAYDRFLPYILMGSLTILTLFLPESF 513
 Db 457 PTFVRLNRVMVCSLDCDIGIITPFTVFRLEVMQALPLILFAVLGLAAGVTLILLPETK 516
 QY 514 GTPLPDTI 521
 Db 517 GDALPETM 524

RESULT 6
 US-09-040-444-2
 ; Sequence 2, Application US/09040444
 ; Patent No. 6063766
 ; GENERAL INFORMATION:
 ; APPLICANT: Koepsell, Hermann
 ; APPLICANT: Grundeman, Dirk
 ; APPLICANT: Gorboulev, Valentin
 ; TITLE OF INVENTION: Transport protein which effects the
 ; TITLE OF INVENTION: Transport of cationic xenobiotics and/or pharmaceuticals,
 ; TITLE OF INVENTION: DNA sequences encoding it and their use.
 ; NUMBER OF SEQUENCES: 6
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner, L.L.P.
 ; STREET: 1300 I Street, N.W., Suite 700
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: USA
 ; ZIP: 20005-3315
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent in Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/040,444
 ; FILING DATE: March 18, 1998
 ; CLASSIFICATION:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: O'Connor, Steven P
 ; REGISTRATION NUMBER: 41,225
 ; REFERENCE/DOCKET NUMBER: 2481.1453-01
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (202)408-4000
 ; TELEFAX: (202)408-4400
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 553 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: peptide
 ; US-09-040-444-2

Query Match 24.8%; Score 714.5; DB 3; Length 553;


```

; Best Local Similarity 35.4%; Pred. No. 8.4e-67;
; Matches 194; Conservative 87; Mismatches 216; Indels 51; Gaps 13;

QY 1 MRDVEYTAFLGEGPQR---LFFELLSASIPNGFTGLSSVFLIATPEHRCRVDPDAAN 57
Db 1 MPTVDDILEQVSGSWFKQAFLLCLLSAFAF---ICVGVFLGTPPDHHCQSPQVAE 57
QY 58 LSS--AMR-----NHTVPLRLRDGREVPHSCRRY-----LATIA-NFSALG 96
Db 58 LSQRCGNSPAEELNVTPEGLPAGAEAFGLQCRRYEVDMNQSAQSCVDPPLASLATNRSHLP 117
QY 97 LEPRDVLGQLESCDGHFESQDVLSTIVTEWNLVCEDDWKAPLTTISLFPVGVLLG 156
Db 118 LGP-----CODGWY--DTPGSSIVTEFNLVCAWSKMLDLFQSLNAGFFFG 162
QY 157 SFISQSLDRFGRKNVLFVTMGOTGFSFLQIFSKNFEMFVFLVGMGOISNYAAVF 216
Db 163 SLGVGFADRGKRLCLLGTVLVNAVSVLMAFSPNYMSMLLFRLLQGLYSKGNWAGYT 222
QY 217 LGTEILGKSVRIIFSTLGVCIYAF--GYMVLPLFAYFIRDWRMLLVALTMPGVLCVALW 274
Db 223 LITEFVSGSR---RTVAIMYQMAFTVGLVALTGLAYALPHRWLQLAVALSLPTFLFLYY 279
QY 275 WFIPESRWLLSOGREFAEVLIRKAANGIVVPSTIFDPSLQDLSSKKQOQSHNILD 334
Db 280 MCVPEPRLWLLSOKRNTAIIKHIAOKNGKLPADKMLSLDEEDVTEK--LSPSFADL 337
QY 335 LRTWIRMTVIMSIMLMWTISVGVFGLSLDTPNLHGDIIVNCFLSAMVEVPAYVLAWL 394
Db 338 FRTPLRKRTFTLMYLFWTFDSVLYOGLLHMGATSGNLYLDLYLSALVEIPGAFIALITI 397
QY 395 QYLPYRYSMATLFLGGSVLLFMQLVPPDLYLATVLMVWKGFGVTAFAFVYVYTAELY 454
Db 398 DRVGRIYMAVSNLLAGAACLV-FISPDHLWLNIIIMCVGRMGITIIAQMICLVNAELY 456
QY 455 PTVVRNMGVGVSTASRLSGSILSPYFV-LGAYDRFLPYFILMGSILTITALTFLPESF 513
Db 457 PTFVRLNLRVWVSSICDGGIITPFIIVFLREVWQALPILFVLLGGLAAGVTLILLPETK 516
QY 514 GTPLPDTI 521
Db 517 GDALPETM 524

RESULT 7
US-08-647-397-2
; Sequence 2, Application US/08647397
; Patent No. 5972702
; GENERAL INFORMATION:
; APPLICANT: Beier, David R.
; TITLE OF INVENTION: OSTEOCLAST TRANSPORTER
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02210
; COMPUTER READABLE FORM:
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/647,397
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Gates, Edward R.
; REGISTRATION NUMBER: 31,616
; REFERENCE/DOCKET NUMBER: B0801/7048

```

```

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-720-3500
; TELEFAX: 617-720-2441
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 537 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-647-397-2

Query Match 22.6%; Score 652.5; DB 2; Length 537;
Best Local Similarity 29.9%; Pred. No. 2.8e-60; Indels 41; Gaps 12;
Matches 167; Conservative 104; Mismatches 246;

QY 4 YDVTAFLEWGPQFQRLIFLFLSASIPNGFTGLSSVFLIATPEHRCRVDPDAANLSSAWR 63
Db 3 FSEILDRVSGMGFFQYLHVTLTLLALPILGIANHNLQIFATTATPDHHCRRPPNASL-EPW- 60
QY 64 NHTVPLRLRDGREVPHSCRRY---RLATIANFSALGLEPRDVLGQLESCDGHFES 120
Db 61 --VPLG-PNGK--PEKCLRFVHLHPNASLPNDTQGATEP-----CLDGWIYN 102
QY 121 QDVYLSITVTEWNLVCEDDWKAPLTTISLFFVGVLLGSFISGQSLDRFGRKNVLFVTMG 180
Db 103 ST--RTIVTEWDLVCGSNKLEMAQSVFMAGILVGGPVFGLSDRFGKRPILTWSYLL 160
QY 181 TGFSEFLQIFSKNFEMFVFLVGMGOISNYAAVFVLTGTEILGKSVRIIFS-TLGVCIY 239
Db 161 AASGSSAAFSPLTVVYIMFPLFCGCSISGISLSTIILNVWVPTSTRAISSTTIGYC--Y 218
QY 240 AFGWVLPFAYFIRDWRMLLVALTMPGVLCVALWFIPEPRLWLLSOGREFAEVLIRK 299
Db 219 TIGQFVLPGIAYAVPQWRWLQLSVAAFFISLSSWVWVPSIRWLVLSKFSALKTLOR 278
QY 300 AARANGVVPSTIFDPSL-----QDLSSKKQOQSHNILDLLRTWIRMTVIMSIMLMWTI 354
Db 279 VAFNGKKEGEKLTVEELKFNLOKDIKSAVK-YGLSDLFVSLIRRVYTFCLSLAWFAT 337
QY 355 SVGYFGLSLDTPNLHGDIIVNCFLSAMVEVPAYVLAWLLOYLPRRYSMATLFLGGSVL 414
Db 338 GFAYISLAMGVVEEFGVNIYILQIFGGVDIPAKFITILSISYLGRRITQGLLILAGVAI 397
QY 415 LFMQVLPDLYLATVLMVWKGFGVTAFAFVYVYTAELYPTVVRNMGVGVSTASRLGS 474
Db 398 LALIFVSSQMQLDRALAVFGKCLSGSFCLFYTSELYPTVLRQTGMGISINWARGVS 457
QY 475 ILSPYFVYLGAIDRFPLPYILMGSILTITALTFLPESFGTLPDITID-----QMLRV 527
Db 458 MIAPLVKITGELQPFIPNVIFWTMTLLGSSAAFPLETLNRPLPETIEDIQDWYQOTKKT 517
QY 528 KGMKHKRTPSHRMLKDG 545
Db 518 KQPEAEKASQTIPLKTG 535

RESULT 8
US-09-031-392-5
; Sequence 5, Application US/09031392
; Patent No. 5942398
; GENERAL INFORMATION:
; APPLICANT: Tartaglia, Louis A.
; APPLICANT: Weng, Xun
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES
; TITLE OF INVENTION: ENCODING GLUTEX AND USES THEREOF
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA

```

ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows95
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/031,392
FILING DATE: 26-FEB-1998
ATTORNEY/AGENT INFORMATION:
NAME: Meiklejohn, Ph.D., Anita L.
REGISTRATION NUMBER: 35,283
REFERENCE/DOCKET NUMBER: 07334/072001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 494 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-031-392-5

Query Match 8.1%; Score 234; DB 2; Length 494;
Best Local Similarity 24.7%; Pred. No. 3.6e-16;
Matches 112; Conservative 85; Mismatches 158; Indels 98; Gaps 23;

```
QY 144 LTSLFVGVLLGSGFSGQLSDRGK-----NVLFVTMGQTGFSLQIFSKNFEMFV 197
      |||:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 65 LSAIFSVGGMIGSFSGVLFVNFRGRNSMLIVNLLAIAGGCLMGFCKI---AESVEMLI 121

QY 198 VLFVLVGM--GOISNVAAAFVLGTEILGKSVRIIFSTL-----GVCIFYAGYMWVLP 248
      |||:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 122 LGRLLIGLFCGLCTGFVPMYI--GEISPTALRGAFGLNQLGIVIGLVAQIFGLKVI-- 177

QY 249 FAYFIRD-WRMILVALTMPGVLCVALWMTIPESPRW-LISQGRFEEAEVIRKAAKANGI 306
      |||:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 178 --LGTEDLWPLLLGFTLPAIIQCAALPCPSPRELLINRKEEKAKELQLRWGTEDV 235

QY 307 VVPSTIFDPSLQDLSSKKQOSH--NILDLRTWNRWMTIMSLMWT-----ISVGY 358
      |||:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 236 AQ-----DIQEMKDESWMSEKQVTVLELFRAPNYRQPIIISIMLQSQLSGINAVFY 290

QY 359 FGLSLDTPNLHGDIFVNCFLSAMVEVPAY-----VLAWLLIQLYLPRTYSMAT 405
      |||:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 291 YSTGI-----FKDAGVQEPVATIGAGVVNTIFTVVSFVLVERAGRR--TLH 335

QY 406 ALFLGG---SVLLFMQLVPPDLY-----YLATVLVMVGKFGVTAAFSMVYVYTAELYP 455
      |||:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 336 LIGLGMAFCSILMTISLLKDNYSWMSFICIGAILVFVAFFEIGPG-PIPWFIVAELFG 394

QY 456 TVVRNMGVGVSS---TASRLGSILSPYFV-YLGAYDRFLPLVILMGSLTILTAITLFL-L 509
      |||:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 395 QGRPRAMAVACSNWTSNLFVGLFPSPATFYLGAY----VFIVTVFLVFWVFFFKV 450

QY 510 PESFGTLPDPTIDQMLRV-----KGMKHKRTP 536
      |||:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 451 PETRGR---TFEITRAPEGVQGTGRGKGP 479
```

RESULT 9

US-09-299-549-5
Sequence 5, Application US/09299549
Patent No. 6136547
GENERAL INFORMATION:
APPLICANT: Tartaglia, Louis A.
APPLICANT: Weng, Xun
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
TITLE OF INVENTION: GLUTEX AND USES THEREOF
NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows95
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/299,549
FILING DATE: 26-APR-1999
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/031,392
FILING DATE: 26-FEB-1998
ATTORNEY/AGENT INFORMATION:
NAME: Meiklejohn, Ph.D., Anita L.
REGISTRATION NUMBER: 35,283
REFERENCE/DOCKET NUMBER: 07334/072002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 494 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-299-549-5

Query Match 8.1%; Score 234; DB 4; Length 494;
Best Local Similarity 24.7%; Pred. No. 3.6e-16;
Matches 112; Conservative 85; Mismatches 158; Indels 98; Gaps 23;

```
QY 144 LTSLFVGVLLGSGFSGQLSDRGK-----NVLFVTMGQTGFSLQIFSKNFEMFV 197
      |||:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 65 LSAIFSVGGMIGSFSGVLFVNFRGRNSMLIVNLLAIAGGCLMGFCKI---AESVEMLI 121

QY 198 VLFVLVGM--GOISNVAAAFVLGTEILGKSVRIIFSTL-----GVCIFYAGYMWVLP 248
      |||:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 122 LGRLLIGLFCGLCTGFVPMYI--GEISPTALRGAFGLNQLGIVIGLVAQIFGLKVI-- 177

QY 249 FAYFIRD-WRMILVALTMPGVLCVALWMTIPESPRW-LISQGRFEEAEVIRKAAKANGI 306
      |||:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 178 --LGTEDLWPLLLGFTLPAIIQCAALPCPSPRELLINRKEEKAKELQLRWGTEDV 235

QY 307 VVPSTIFDPSLQDLSSKKQOSH--NILDLRTWNRWMTIMSLMWT-----ISVGY 358
      |||:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 236 AQ-----DIQEMKDESWMSEKQVTVLELFRAPNYRQPIIISIMLQSQLSGINAVFY 290

QY 359 FGLSLDTPNLHGDIFVNCFLSAMVEVPAY-----VLAWLLIQLYLPRTYSMAT 405
      |||:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 291 YSTGI-----FKDAGVQEPVATIGAGVVNTIFTVVSFVLVERAGRR--TLH 335

QY 406 ALFLGG---SVLLFMQLVPPDLY-----YLATVLVMVGKFGVTAAFSMVYVYTAELYP 455
      |||:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 336 LIGLGMAFCSILMTISLLKDNYSWMSFICIGAILVFVAFFEIGPG-PIPWFIVAELFG 394

QY 456 TVVRNMGVGVSS---TASRLGSILSPYFV-YLGAYDRFLPLVILMGSLTILTAITLFL-L 509
      |||:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 395 QGRPRAMAVACSNWTSNLFVGLFPSPATFYLGAY----VFIVTVFLVFWVFFFKV 450

QY 510 PESFGTLPDPTIDQMLRV-----KGMKHKRTP 536
      |||:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 451 PETRGR---TFEITRAPEGVQGTGRGKGP 479
```

RESULT 10


```

;
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 493 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-031-392-10

Query Match          7.0%; Score 208; DB 5; Length 492;
Best Local Similarity 22.2%; Pred No. 2e-13;
Matches 108; Conservative 87; Mismatches 165; Indels 126; Gaps 22;

QY 119 FSDVYLSTIVTENNIVCEDDKAPITISLFFVGLVLSFISQGLSDRGRK-----NV 172
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 52 YGESILPTTLTLTWS-----LSVAIFS VGGIGSFVGLFVNFRGRNMLMNL 101
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 173 LFVTMGOTGFSFLOIFSKNFEMFVFLVGM--COISNYVAAFVLGTEILGKSVRIIF 230
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 102 LAFVSAVLGMGFSKL---GKSFEMLIILGRFIIGYGLTTCFPMYV--GEVSPATFRGAL 156
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 231 STL-----GVCIFYAFGYMVLPLFAYFIRD-----WRMLLVALTMPGVLCVALWM 275
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 157 GTLHQLGIVVGIILIAQVFL-----DSIMGNKDLWFLLSIIFIPALLOCIVIP 205
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 276 FIPESRW-LISGRPEEAIVIRKAAKANGIVPSTIPDPSELQDLSKKQOSH----- 329
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 206 FCPESPRFLINRNEENRAKSVLKLKRGTDVTV-----HDLOEMKEESQMMRE 254
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 330 ---NILDLRTWNIRMTTMSIMLMT-----ISVGYEGLSLDTPNLHGDIFVNCFLSA 380
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 255 KVTIIELEFRSPAYROPILIAVVQLSQOOLSGINAVFYSTSI-----FEKA 301
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 381 MVEVPAY-----VLAWLLLOLYLPRR--YSMATALFLGSGVLLFMOLV---P 421
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 302 GVQOPVYATIGSIVNTAFTVVSFLFVVERAGRRTLHLIGLAGMACQAIIIMTIALALLQL 361
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 422 PDLYIYATVLMVVGKFGVTAAFS-----MVYYTAEIYPTVVRNMGVSVSTASRLGS-I 475
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 362 PMWSYLSIAI---FGFAFFEVGPGPIPFVAELESQGRPAIAIAGVFSNNTSNI 417
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 476 LSPFYFYL---GAYDRFLPYILMGLSLTILTAILFL-LPESFGTPLDPTIDQMLRVKGM 530
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 418 VGMCFQYVQLCGPY-----VFIIFFVLLVLFIRYFKVPETKGRFTDEIASGFRQGGAS 473
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 531 KHKRTP 536
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 474 QSDKTP 479

RESULT 12
US-09-031-392-10
; Sequence 10, Application US/09031392
; Patent No. 5942398
; GENERAL INFORMATION:
; APPLICANT: Tartaglia, Louis A.
; APPLICANT: Weng, Xun
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES
; TITLE OF INVENTION: ENCODING GLUTEX AND USES THEREOF
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/031.392
; FILING DATE: 26-FEB-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Meikiejohn, Ph.D., Anita L.
; REGISTRATION NUMBER: 35,283
; REFERENCE/DOCKET NUMBER: 07334/072001
; TELECOMMUNICATION INFORMATION:

```

```

;
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 493 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-031-392-10

Query Match          7.0%; Score 203; DB 2; Length 493;
Best Local Similarity 25.0%; Pred No. 6.8e-13;
Matches 111; Conservative 76; Mismatches 147; Indels 110; Gaps 21;

QY 144 LTISLFFVGLVLSFISQGLSDRGRK-----NVLFVTMGOTGFSFLOIFSKNFEMFV 197
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 68 LSVSIFAVGGMIGSFLVGXIGNLGRKXKAMLVNVLAIAGLLMG---LAKXAXSPEMLI 124
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 198 VLFVLVGM-----GOISNYVAAFVLGTEILGKSVRIIFSTLGCVCIFYAFGYM 244
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 125 LGRFIIGLYCGLSSGVPMYVGEISPTALRGALGT-----LNQLGIVIGILIA-Q 173
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 245 VLPFLFAYFIRD--WRMLLVALTMPGVLCVALWMFIPESPRWLI-----SQGRPEEAIVIR 298
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 174 VLGLDSLILGNESLWPLLLGLTGVPAQLQLLLPCPSPRYLLINKNEEARAKKALQRLR 233
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 299 KAAKANGIVPSTIPDPSELQD-----LSSKKQOSHNLIDLLRTWNIRMTTMSIMLMT- 353
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 234 GTADVQSEV-----AEMKDESXMXSEKXVS--VLELFRSRXYRQPVIIAIVLQLSQ 283
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 354 ----ISVGYEGLSLDTPNLHGDIFVNCFLSAMVEVPAY-----VLAWLLLO 395
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 284 QLSGINAVFYSTSI-----FEKAGVGQPVYATIGAGVNVNVTVVSVFVVE 330
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 396 YLPRYSMATALFVG-----SVLLFMQLVPPDLYLATVLMVYGKFGVTAAFS-----MV 446
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 331 RAGRR--TLHLGLGGMAGCAVLMTIALALDQVPMWSYVSIIVAFGFVAFVFGPGPIP 388
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 447 VVYTAELYPTVVRNMGVGVSS-----TASRLGSLTSPYFV-YLGAYDRFLPYILMGLSITIL 501
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 389 WFIVAELEFSQGRPAIAIAGVAGFSNNTSNTFVGLLFQYIAELLGPY----VFIVFAVLLLL 444
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 502 TAILT-LFLPESFGTPLDPTIDQM 524
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 445 FFITFLKVPETKGR-----TFDEI 464
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 13
US-09-299-549-10
; Sequence 10, Application US/09299549
; Patent No. 6136547
; GENERAL INFORMATION:
; APPLICANT: Tartaglia, Louis A.
; APPLICANT: Weng, Xun
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
; TITLE OF INVENTION: GLUTEX AND USES THEREOF
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/299,549

```

```

; FILING DATE: 26-APR-1999
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/031,392
; FILING DATE: 26-FEB-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Meiklejohn, Ph.D., Anita L.
; REGISTRATION NUMBER: 35,283
; REFERENCE/DOCKET NUMBER: 07334/072002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 493 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-299-549-10

Query Match 7.0%; Score 203; DB 4; Length 493;
Best Local Similarity 25.0%; Pred. No. 6.8e-13;
Matches 111; Conservative 76; Mismatches 147; Indels 110; Gaps 21;

QY 144 LTISLFFVGLGSGISQSDRGRK-----NVLFVTMGQTGFSFLQIFSKNFEMFV 197
DB 68 LSVSIFAVGGMIGSFLVXIGNLGRKXAMLVNVLAIAGLLMG---LAKXAXSEFMILI 124
QY 198 VLFVLGVM-----GQISNVAAFAVLGTGLGKSVRIIFSTGLGVCIFVAFGYM 244
DB 125 LGRFTIIGLYCGLSGVPMYVGEISPALRGALGT-----LNQLGIVIGILIA-Q 173
QY 245 VLPFLFAYFIRO--WRMLIVALTMPGLCVLAWMFIPESPRWLI-----SQRFEAEVFIIR 298
DB 174 VLGLDSLGNESLPLLLGLTGVPALLQLLLPCCPESPRYLLINKNEEARAKKALQRLR 233
QY 299 KAAKANGIVPSTIDFSELOD-----LSSKKQOSHNLDLRTWNRMYTINSIMLWMT- 353
DB 234 GTADYSOEV-----AEMKDESRRXMXSEKXVS--VLELFSRXRYQPVIIAIVLQLSQ 283
QY 354 -----ISVGYGLSLDTPNLHGDFVNCFLSAMVEVPAY-----VLAWLLQLQ 395
DB 284 QLSGINAVFYVTSI-----FEKAGVGQPVYATIGAGVYNTVTVSVFVVE 330
QY 396 YLPRYSMATALLFGG-----SVLLFMQLVPPDLYLATVLMVGKFGVTAAFS-----WV 446
DB 331 RAGRR--TLHLGLGGMAGCAVLMTIALALDDQVPMMSYVSIVAIFGVFAFFEVGPGPIP 388
QY 447 YVYTAELPTVVRNMGVGVSS-----TASRLGSIILSPYFV-YLGAYDRFLPYILMGLTIL 501
DB 389 WFIVAELFSQGRPAIAIAGVGSFNWTSNFIVGLLFQXIAELLGPY-----VFIVFVALLLL 444
QY 502 TAILT-LFLPESFGTLPDPTIDQM 524
DB 445 FIFTFLKVPETKGR---TFDEI 464

RESULT 14
US-08-647-397-4
; Sequence 4, Application US/08647397
; Patent No. 5972702
; GENERAL INFORMATION:
; APPLICANT: Beier, David R.
; APPLICANT: Brady, Kevin P.
; TITLE OF INVENTION: OSTEOCLAST TRANSPORTER
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: MA
; COUNTRY: USA

```

```

; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/647,397
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Gates, Edward R.
; REGISTRATION NUMBER: 31,616
; REFERENCE/DOCKET NUMBER: B0801/7048
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-720-3500
; TELEFAX: 617-720-2441
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 109 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; US-08-647-397-4

Query Match 6.9%; Score 198; DB 2; Length 109;
Best Local Similarity 40.8%; Pred. No. 2.2e-13;
Matches 42; Conservative 22; Mismatches 39; Indels 0; Gaps 0;

QY 392 LLLYLPRLRYSMATALLFGSVLLFMQLVPPDLYLATVLMVGKFGVTAAFSMVYVYTA 451
DB 6 LSLYLGHTTQAAALLAGGAILALTFPLDLQTVTLAVLFGKLSLSSFSCLFLYTS 65
QY 452 ELYPTVVRNMGVGVSSVTSRLGSIILSPYFVYLGAYDRFLPYIL 494
DB 66 ELYPTVIRQTGMGVSNLMTVRGVMSVXLKITGEVQFPIFNII 108

RESULT 15
US-09-031-392-6
; Sequence 6, Application US/09031392
; Patent No. 5942398
; GENERAL INFORMATION:
; APPLICANT: Tartaglia, Louis A.
; APPLICANT: Weng, Xun
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/031,392
; FILING DATE: 26-FEB-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Meiklejohn, Ph.D., Anita L.
; REGISTRATION NUMBER: 35,283

```

REFERENCE/DOCKET NUMBER: 07334/072001
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 617/542-5070
 TELEFAX: 617/542-8906
 TELEX: 200154
 INFORMATION FOR SEQ ID NO: 6:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 509 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-09-031-392-6

Query Match 6.7%; Score 193; DB 2; Length 509;
 Best Local Similarity 24.1%; Pred. No. 8.1e-12;
 Matches 125; Conservative 93; Mismatches 175; Indels 126; Gaps 28;

QY 89 IANFSAL--GLEPGRVDL-----CQLQESCLDQW-----EFSQDVLYSTIVTENNLYC 136
 DB 26 LAVFSAVLGLOFYNGIVINAPQVIEQSYNETWLGROGPEGSPIPPGLTTLW-----81
 QY 137 EDDKAPLITISLFYVGLSGLSQSDRFGRK-----NVLFVTMGMTQSFSLQIFS 190
 DB 82 -----ALSVAIFSGVGMISSEFLGIISQWLGKRAMLVNVLAVLGGSLMG---LANAA 132
 QY 191 KNFEMFVFLVVG--MQQISNYVAAFVLGTEILGKSVRIIFSTL-----GVCIFYAF 241
 DB 133 ASYEMLIILGRELIGAYSGLSGLVPMV--GETAPHLRGALGTNLQLAIVIGILIAQVL 190
 QY 242 GYMVLPLFAYETIRWRMLLVALTMPGVLCAVMWFIPESPRWLISQGRFEEAEVIRKAA 301
 DB 191 GLESLGTASL--WPLLGLTLVLPALLQLVLLPFCPEPRYLYI---IQNLGSPARKSL 244
 QY 302 K-----ANGIVPSTIFDPSELDLSSK--KQOSHNIIDLRTWNIRMTMSIMLWM 352
 DB 245 KRLTGWADVSGVL-----AELKDKRKLREPLSLQLLQIGSRTHRQPLIAVVLQ 296
 QY 353 T-----ISVGYFGLSLDTPNLHGDI FVNCFLSAMVEVPAY-----VLAWLL 393
 DB 297 SQQLSGINAVFYYSI-----FETAGVGPAYATIGAGVNVTVTLVSLL 343
 QY 394 LQYLPFR--YSMATALFLGGSVLLEMLV-----PPDLYLATVLMVVGKFGVTAAPS---444
 DB 344 VERAGRRTLHLGLAGMCGCAILMTVALLLRLVPAMSYYSIVAI---FGFVAFPEIGP 399
 QY 445 --MVVYVTAELYPTVVRNMGVGSSTASRLGS--ILSPYFVLGAYDRFLPYI--LMGSLTI 500
 DB 400 GPDPWFIVAELESQCPRAANAVAGFSNWTNFTIIGMGFOYVA--EAMGPYVLLFAVLL 457
 QY 501 LTAILTFL--PESFGTLPDPTIDOMLRVGMKHKRTPS 537
 DB 458 LGFFITFLRPVETGR---TFDQ---ISAAFHR--TPS 488

Search completed: August 16, 2001, 13:58:54
 Job time: 216 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 16, 2001, 13:59:27 ; Search time 28.45 Seconds
(without alignments)
1491.361 Million cell updates/sec

Title: US-09-521-195-3

Perfect score: 2883

Sequence: 1 MRDYDEVTAFLGEGWPFQRL.....HTRMLKDGQRPRTILKSTAF 557

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_68: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2883	100.0	557	2 JW0089	organic cation tra
2	2496	86.6	557	2 JE0346	high-affinity carn
3	777	27.0	576	2 T22509	hypothetical prote
4	733	25.4	593	2 JC4884	organic cation tra
5	712	24.7	556	2 S50862	organic cation tra
6	589	20.4	794	2 T27870	hypothetical prote
7	547	19.0	745	2 T16565	hypothetical prote
8	522	18.1	527	2 T01019	transport protein
9	483	16.8	539	2 C96758	probable protein
10	470	16.3	518	2 B86299	hypothetical prote
11	411	14.3	528	2 T21682	hypothetical prote
12	382	13.3	540	2 T25851	hypothetical prote
13	380	13.2	515	2 B96825	hypothetical prote
14	368	12.8	521	2 H86298	hypothetical prote
15	355	12.3	529	2 T23190	hypothetical prote
16	341	11.8	454	2 F75580	probable sugar tra
17	316	11.0	455	2 B83213	probable MFS trans
18	308	10.7	469	2 B65058	hypothetical prote
19	298	10.4	724	2 I50531	transmembrane tran
20	295	10.2	423	2 S74046	probable sugar tra
21	291	10.1	435	2 T15290	hypothetical prote
22	287	10.0	400	2 C69757	transporter homolo
23	287	10.0	422	2 G72234	hypothetical prote
24	279	9.7	524	2 T27082	hypothetical prote
25	274	9.5	683	2 S34961	synaptic vesicle p
26	272	9.4	478	2 T33985	hypothetical prote
27	271	9.4	480	2 T23608	hypothetical prote
28	267	9.3	443	2 E64725	yaaU protein - Esc
29	267	9.3	443	2 H85485	probable transport

ALIGNMENTS

RESULT 1

JW0089

organic cation transporter protein 2 - human

N:Alternate names: OCTN2

C:Species: Homo sapiens (man)

C>Date: 18-Jun-1998 #sequence_revision 10-Jul-1998 #text_change 28-May-1999

C:Accession: JW0089

R:Wu, X.; Brasad, P.D.; Leibach, F.H.; Ganapathy, V.

Biochem. Biophys. Res. Commun. 246, 589-595, 1998

A:Title: cDNA sequence, transport function, and genomic organization of human OCTN2,

A:Reference number: JW0089; MUID:98289574

A:Accession: JW0089

A:Molecule type: mRNA

A:Residues: 1-557 <WUA>

A:Cross-references: GB:AF057164; NID:g3273740; PIDN:AAC24828.1; PID:g3273741

A:Experimental source: placenta

C:Comment: This transporter functions in the elimination of cationic drugs and other

Query Match 100.0%; Score 2883; DB 2; Length 557;
Best Local Similarity 100.0%; Pred. No. 7.8e-216;
Matches 557; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MRDYDEVTAFLGEGWPFQRLIFLLSASTIIPNGFTGLSSVFLIATPEHRCRVPDAANLSS	60
Db	1	MRDYDEVTAFLGEGWPFQRLIFLLSASTIIPNGFTGLSSVFLIATPEHRCRVPDAANLSS	60
Qy	61	AWRNHTVPLRLRDGREVPHSCRRYRLATIANFSALGLEPGRDVLGQLEQESCLDGWEFS	120
Db	61	AWRNHTVPLRLRDGREVPHSCRRYRLATIANFSALGLEPGRDVLGQLEQESCLDGWEFS	120
Qy	121	QDVYLSTIVTEWNLVCEDDWKAPLISLFFVGVVLGSGFISGQLSDRFGKKNVLFVTMGQ	180
Db	121	QDVYLSTIVTEWNLVCEDDWKAPLISLFFVGVVLGSGFISGQLSDRFGKKNVLFVTMGQ	180
Qy	181	TGFSFLQIFSKNFEMFVFLVGVGMQISNYVAARVLTGLGKSVRIIFSTLGVCI EYA	240
Db	181	TGFSFLQIFSKNFEMFVFLVGVGMQISNYVAARVLTGLGKSVRIIFSTLGVCI EYA	240
Qy	241	FGYVLPPLFAYFIRDWRMLLVATMPGVLCVAMWFIPESPRWLISQGRFEAEVIRKA	300
Db	241	FGYVLPPLFAYFIRDWRMLLVATMPGVLCVAMWFIPESPRWLISQGRFEAEVIRKA	300
Qy	301	AKANGIVVPSTIFDPSQLSSKKQKQSHNILDLLRTWNIRMTIMSLMNTISVGVFG	360
Db	301	AKANGIVVPSTIFDPSQLSSKKQKQSHNILDLLRTWNIRMTIMSLMNTISVGVFG	360
Qy	361	LSLDPNLHGDIFVNCFLSAMVEVPAYVLAWLLQLYLPRRYSMATALFLGGSVLLFMQLV	420
Db	361	LSLDPNLHGDIFVNCFLSAMVEVPAYVLAWLLQLYLPRRYSMATALFLGGSVLLFMQLV	420
Qy	421	PPDLYLATVLVMGKFGVTAAFSMVYVVTABLYPTVVRNMGVGSSTASRLGSLSPYF	480

hypothetical prote
hypothetical prote
probable transport
probable MFS trans
synaptic vesicle p
metabolite transpo
hypothetical metab
hypothetical prote
arabinose-proton s
hypothetical prote
sugar transporter
probable transport
metabolite transpo
hypothetical prote
metabolite transpo

Db 421 PPDLVYLATVLMVGVKFGVTAFAFVSVVYTAELYPVVRNMGVSVSTASRLGSLSPYF 480
QY 481 VYLGAYDRFLPYLLMGSLTILTAITLFLPESFGTLPDIDOMLRVKMGKHKRTPSHTR 540
Db 481 VYLGAYDRFLPYLLMGSLTILTAITLFLPESFGTLPDIDOMLRVKMGKHKRTPSHTR 540
QY 541 MLKDGQERTILKSTAF 557
Db 541 MLKDGQERTILKSTAF 557

RESULT 2
JE0346
high-affinity carnitine transporter, CTL - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
C:Accession: JE0346
R: Sekine, T.; Kusuhabara, H.; Utsunomiya-Tate, N.; Tsuda, M.; Sugiyama, Y.; Kanai, Y.; Endo
Biochem. Biophys. Res. Commun. 251, 586-591, 1998
A:Title: Molecular cloning and characterization of high-affinity carnitine transporter f
A:Reference number: JE0346; MUID:99011422
A:Accession: JE0346
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-557 <SEK>
A:Cross-references: DDBJ: AB017260; NID: g3869208; PIDN: BAA34399.1; PID: g3869209

Query Match 86.6%; Score 2496; DB 2; Length 557;
Best Local Similarity 85.3%; Pred. No. 7.8e-186;
Matches 475; Conservative 42; Mismatches 40; Indels 0; Gaps 0;

QY 1 MRDYDEVTAFLGEMGPFQRLIFLLSASIIPIGFTGLSSVFLIATPEHRCRVPDAANLSS 60
Db 1 MRDYDEVTAFLGEMGPFQRLIFLLSASIIPIGFTGLSSVFLIATPEHRCRVPDAANLSS 60
QY 61 AWNHVPLRLDGRVPHSCRRYRLATIANFSLGLEPRDVLGLOEQESCLDGEWFS 120
Db 61 AWNHVPLRLDGRVPHSCRRYRLATIANFSLGLEPRDVLGLOEQESCLDGEWFS 120
QY 121 QYVLSITVTENLVNCEDDWKAPLITISFFVGVLLGSFISGOLSDRGKKNVLFVTMGQ 180
Db 121 KOVFLSTIVTENDLVCKDKWKAPLITISFFVGVLLGSFISGOLSDRGKKNVLFVTMGQ 180
QY 181 TGFSLQIFSKNFEMFVFLVGVGQISNYVAAFVLGTEILGKSVRIIFSLGVCIFYA 240
Db 181 TGFSLQIFSKNFEMFVFLVGVGQISNYVAAFVLGTEILGKSVRIIFSLGVCIFYA 240
QY 241 FGVMVPLPAYFIRDMRLMLLVALTPGVLCVALLWFPESRWLISQGRFEAEVIRKA 300
Db 241 FGVMVPLPAYFIRDMRLMLLVALTPGVLCVALLWFPESRWLISQGRFEAEVIRKA 300
QY 301 AKANGIVPSTIFDPSELQDLSKKQSHHIIYDLVTRNIRIITIMSLWLTISVGYFG 360
Db 301 AKANGIVPSTIFDPSELQDLSKKQSHHIIYDLVTRNIRIITIMSLWLTISVGYFG 360
QY 361 LSLDTPNLHGDIFVNCFLSAMVEVPAYVLAALLQYLPRRYSMATLFLGGSVLLF 420
Db 361 LSLDTPNLHGDIFVNCFLSAMVEVPAYVLAALLQYLPRRYSMATLFLGGSVLLF 420
QY 421 PPDLVYLATVLMVGVKFGVTAFAFVSVVYTAELYPVVRNMGVSVSTASRLGSLSPYF 480
Db 421 PSELVYSLATVLMVGVKFGVTAFAFVSVVYTAELYPVVRNMGVSVSTASRLGSLSPYF 480
QY 481 VYLGAYDRFLPYLLMGSLTILTAITLFLPESFGTLPDIDOMLRVKMGKHKRTPSHTR 540
Db 481 VYLGAYDRFLPYLLMGSLTILTAITLFLPESFGTLPDIDOMLRVKMGKHKRTPSHTR 540
QY 541 MLKDGQERTILKSTAF 557
Db 541 TOKDGESPTVLKSTAF 557

RESULT 3
T22509
hypotheical protein F52F12.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T22509
R: Matthews, L.
submitted to the EMBL Data Library, December 1996
A:Reference number: Z19573
A:Accession: T22509
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-576 <WIL>
A:Cross-references: EMBL: Z83228; PIDN: CAB05732.1; GSPDB: GN00019; CESP: F52F12.1
A:Experimental source: clone F52F12
C:Genetics:
A:Gene: CESP: F52F12.1
A:Map position: 1
A:Introns: 16/1; 23/3; 50/3; 80/3; 108/2; 134/3; 221/1; 257/2; 319/3; 358/2; 490/3; 5

Query Match 27.0%; Score 777; DB 2; Length 576;
Best Local Similarity 31.4%; Pred. No. 1.5e-52;
Matches 165; Conservative 118; Mismatches 225; Indels 18; Gaps 6;

QY 3 DYDEVTAFLGEMGPFQRLIFLLS-ASIIPIGFTGLSSVFLIATPEHRCRVPDAANLSSA 61
Db 16 DFDFVLEQVNGYQIVFFIICLPTLSAFSAFNPVGVGNPHPTCHIPGKEYLRP 75
QY 62 WRNTHVPLRLDGRVPHSCRRYRLATIANFSLGLEPRDVLGLOEQESCLDGEWFSQ 121
Db 76 LTNDTQIL-----SCKQXNETQINVFRAFTSAP-VDTYSDRISLVPCCQNGWDYN 124
QY 122 DVYLSITVTENLVNCEDDWKAPLITISFFVGVLLGSFISGOLSDRGKKNVLFVTMGQ 181
Db 125 STYLDLSLTFEFLNVCDDQAWIEIISFFYVGSFIGNCLFGYVADKFGRRSRFFVILT 184
QY 182 GFSFLQIFSKNFEMFVFLVGVGQISNYVAAFVLGTEILGKSVRIIFSLGVCIFYA 241
Db 185 VCGTASSFAKDIESFIILRFTGLAPALFQIFPICMEFPMGNSGR-IFSLGTLSTLFFGA 243
QY 242 GYVPLPAYFIRDMRLMLLVALTPGVLCVALLWFPESRWLISQGRFEAEVIRKAA 301
Db 244 AMALLGVVAFIRRMQLTFFCNAPFAFYIYFFELPESRWSVSVGKNADAKKLLKIA 303
QY 302 KANG---IVVPSTIFDPSELQDLSKKQ--OSHNIIDLRTNIRMTIMSLWLTISV 356
Db 304 KMNGKSNVDDELVDLSMKHQAEEKETRSNVTDLFTPNLRKTLTIIVYVWNAI 363
QY 357 GYFGLSDTPNLHGDIFVNCFLSAMVEVPAYVLAALLQYLPRRYSMATLFLGGSVLLF 416
Db 364 IYNGLTNLVSNLPPVDDYWSFIINGAVELPGYFVVMPLLCACAGRRWTLAATMIVCGIGCVS 423
QY 417 MQLVPPDLYLATVLMVGVKFGVTAFAFVSVVYTAELYPVVRNMGVSVSTASRLGSL 476
Db 424 AMFMDGYPMLVASASTFGKFGVSGFAVYIIFAGELYPTVVRNMGVSVSGVAGSGLLL 483
QY 477 SPYFVYLGAYDRFLPYLLMGSLTILTAITLFLPESFGTLPDID 522
Db 484 APHTVNLGKIVKILPLLIMGLMALSAGILFTFLPETIGAPLPMTIE 529

RESULT 4
JC4884
organic cation transporter protein 2 - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 10-Sep-1996 #sequence_revision 18-Oct-1996 #text_change 05-Nov-1999
C:Accession: JC4884
R: Okuda, M.; Saito, H.; Urakami, Y.; Takano, M.; Inui, K.
Biochem. Biophys. Res. Commun. 224, 500-507, 1996
A:Title: cDNA cloning and functional expression of a novel rat kidney organic cation
A:Reference number: JC4884; MUID:96295517

A:Accession: JC4884

A:Molecule type: mRNA

A:Residues: 1-593 <OKU>

A:Cross-references: DBJ:D83044; NID:gl502282; PIDN:BAAL1754.1; PID:d1012421; PID:gl5022

A:Experimental source: kidney

C:Comment: This protein is responsible for the transport of cationic drugs in kidney.

Query Match 25.4%; Score 733; DB 2; Length 593;
Best Local Similarity 35.2%; Pred. No. 3.9e-49;
Matches 198; Conservative 89; Mismatches 234; Indels 42; Gaps 15;

QY 1 MRDYDEVTAFLGEGWGFQRLIFLLSASITPNFTGL--SSVFLATPEHRCRVPDAANL 58

Db 1 MSTVDDILEHGFHFLQKQTFELL--ALLSCAFPIYGVIFLFTPDHHCWSGAACL 58

QY 59 SS--AWR-----NHTVP-LRLRDGVRPHSCRRYR-----LATANFSAFLGLEPGRDV 103

Db 59 SQRCSQAEELNYTPVGLGSPDEASFLSQCMRYEDWNQSTLDCVPLSLA-----A 112

QY 104 DLQLEQESCLDGEFSQDVYLSITVTENLVCEDDWKAPLISLFFVGVLLGSFISQL 163

Db 113 DRNQLPLGCEHGWYNTFEG--SSIVTEFNLCASWMLDLFQSVVNVGFFIGAMMIGYL 170

QY 164 SDRFGKNVLFVTMGMTGFSFLQIFSKNFEMFVVLVFLVGMGQISNVAAFLVGTETLG 223

Db 171 ADRFGKFCLLVTLINATISGALMAISPNAWLVFRFLOGLVSRAGWLGILITFVVG 230

QY 224 KSVRIIFSTLGVCIYAF--GYMVLPLFAFYIRDRWMLLVALTMPLGVALWFWFIPSP 281

Db 231 LGYR--RWGICYAFTVGLLILAGVAVIPNWRWLFQAVTLNFCFLYFWCIPESP 287

QY 282 RWLISGRPEAEVIRKAANGANGIVVPSTIEDPSLQDLSKKQOOSHNLILLRTWNR 341

Db 288 RWLISQNKIVKAKIITKIAKKNGKSPVLSQNLTPDEDAKKLKP--ILDLVTRTPQR 345

QY 342 MVTIMSLMWTISVGYFLSLDTPNLHGD-IFVNCFLSAMVEVPAYVLAWLLOYLPRR 400

Db 346 KHTLIMYNFTSSVLYQGLIMHM-GLAGNIYLDFFYSALVEFPAAFIITIDRVGR 404

QY 401 YSMATLFGGSVLLFMQLYLAIVLVLMVGRFGVTAAFSVYVYVYVYVYVYVYVYVYV 460

Db 405 YPNAVSNVAGAACLASVFPDPLQWLKITIACLRGMGITMAYEMVCLVNAELYPIYRN 464

QY 461 MGCVSVSTASRLSGISLSPFYV-LGAYDRFLPYLMGSLTILTAITLPLPSFGTLPD 519

Db 465 LGVLVCSMDIGGIITPFLVRLTDIWMFPLVFAVGLVAGALVLLLPETKGLPE 524

QY 520 TID-----QMLRVKGMKHKRTPS 537

Db 525 TIEDAENKQPRKERNLPPS 547

RESULT 5
S50862
organic cation transport protein Ocr1 - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 05-Nov-1999
C:Accession: S50862; S78533; I58089
R:Gruendemann, D.; Gorboulev, V.; Gambaryan, S.; Veyhl, M.; Koepsell, H.
Nature 372, 549-552, 1994
A:Title: Drug excretion mediated by a new prototype of polyspecific transporter.
A:Reference number: I58089; MUID:95082907
A:Accession: S50862
A:Molecule type: mRNA
A:Residues: 1-556 <GRU>
A:Cross-references: EMBL:X78855; NID:g633621; PIDN:CAA55411.1; PID:g633622
A:Experimental source: kidney
R:Gorboulev, V.G.
submitted to the EMBL Data Library, January 1995
A:Reference number: S78533
A:Accession: S78533
A:Molecule type: mRNA

A:Residues: 1-342, 'N', 344-556 <GOR>

A:Cross-references: EMBL:X78855; NID:g633621; PIDN:CAA55411.1; PID:g633622

C:Keywords: glycoprotein; phosphoprotein; transmembrane protein

F:20-46/Domain: transmembrane #status predicted <TM1>

F:154-171/Domain: transmembrane #status predicted <TM2>

F:178-197/Domain: transmembrane #status predicted <TM3>

F:243-260/Domain: transmembrane #status predicted <TM4>

F:267-283/Domain: transmembrane #status predicted <TM5>

F:350-366/Domain: transmembrane #status predicted <TM6>

F:380-398/Domain: transmembrane #status predicted <TM7>

F:406-425/Domain: transmembrane #status predicted <TM8>

F:435-452/Domain: transmembrane #status predicted <TM9>

F:469-485/Domain: transmembrane #status predicted <TM10>

F:494-514/Domain: transmembrane #status predicted <TM11>

F:71,97,113,432/Binding site: carbohydrate (Asp) (covalent) #status predicted

F:286,292/Binding site: phosphate (Ser) (covalent) (by protein kinase C) #status pred

F:296,343,550/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status

Query Match 24.7%; Score 712; DB 2; Length 556;

Best Local Similarity 33.8%; Pred. No. 1.5e-47;

Matches 193; Conservative 96; Mismatches 230; Indels 52; Gaps 14;

QY 1 MRDYDEVTAFLGEGWGFQRLIFLL--SASITPNFTGLSSVFLIATPEHRCRVPDAAN 57

Db 1 MPTVDDVLEQVGEGWGFQKQAFLLCLLISASLAPI-VVGI--VFLGTPGHYQCNPGVAE 57

QY 58 LSS--AWR-----NHTVP-LRLRDGVRPHSCRRYL-----ATTANFSA 95

Db 58 LSQRCSQAEELNYTPVGLGSPDEASFLSQCMRYEDWNQSTLDCVPLSLVANSRL 117

QY 96 GLEPCRDVLDLQLEQESCLDGEFSQDVYLSITVTENLVCEDDWKAPLITSLFVGVLL 155

Db 118 PLGP-----CEHGWY--DTPGSSIVTEFNLCVGDWAKVLDLQSCVNLGFFL 162

QY 156 GSFISGQLSDRFGKRVNLFVTMGMTGFSFLQIFSKNFEMFVVLVFLVGMGQISNVYAA 215

Db 163 GSVLVGYIADRFGRKLCCLLVTLVTSVSGVLTAVADYTSMLLFRLLQGMVSKGWSVGY 222

QY 216 VLGTEILKSVRIIFSTLGVCIYAF--GYMVLPLFAFYIRDRWMLLVALTMPLGVAL 273

Db 223 TLITFVSGSYR--RTTAILYQMAFTVGLVGLAGVAYAIIPDRWMLQVLAISLFTFLFLY 279

QY 274 WVFIPESPRLLISQGRPEEAEEVIRKAANGANGIVVPSTIFDPSELQDLSKKQOOSHNL 333

Db 280 YWFPESPRLLISQGRTRAVRIMEQIAQKNGKVPADLKMCLCEADASEKR--SPSFAD 337

QY 334 LLRTWNIRMTIMSLMWTISVGYFLSLDTPNLHGDIFVNCFLSAMVEVPAYVLAWL 393

Db 338 LFRTPTLRKHVTILMYLWFSCAVLYQGLIMHVGGATGANLYLDFVSSILVEPFAAFILVT 397

QY 394 LOYLPRRYSMATALELGGSVLLFMQLVPPDLYLYLATVLMVKGKGVTAFAFVYVYVYTAEL 453

Db 398 IDRIGRYPIAASNLVITGAACLMIFIPHELHNLNVLACLRMGATVILQVCLVNAEL 457

QY 454 YPTVVRNMGVSVSTASRLSGISLSPFYV-LGAYDRFLPYLMGSLTILTAITLFLPES 512

Db 458 YPTFTRNLGMVCSALDGLGIFTFPFWFRMLMEVQWALPLILFGLVGLTAGATMILLPET 517

QY 513 FGCTPLPDTIDQMLRVYKGMKHKRTPSHRMLK 543

Db 518 KGVALPETIEEAENL-GRKSKAKENTIVLQ 547

RESULT 6

T27870

hypothetical protein ZK455.8 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000

C:Accession: T27870

R:White, S.

submitted to the EMBL Data Library, November 1995

A:Reference number: Z20432

A:Accession: T27870
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-794 <WIL>
A:Cross-references: EMBL:Z66567; PIDN:CAA91492.2; G
A:Experimental source: clone ZK455
C:Genetics:
A:Gene: CRSP:ZK455.8
A:Map position: X
A:Introns: 52/3; 119/2; 158/1; 207/3; 291/3; 335/2;

[illegible]

```

RESULT      7
T15656     hypothetical protein K05F1.6 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #tex
C:Accession: T15656
R:Wohldmann, P.
submitted to the EMBL Data Library, June 1995
A:Description: The sequence of C. elegans cosmid K05F1.
A:Reference number: Z18537
A:Accession: T15656
A:Status: preliminary; translated from GB/EMBL/DBJ

```

A: Molecule type: DNA
A: Residues: 1-745 <WOH>
A: Cross-references: EMBL:U93977; NID:g868173; PTD:g868176; PIDN:AAA68713.1; CESP:K05F1.6
A: Experimental source: strain Bristol N2
C: Genetics:
A: Gene: CESP:K05F1.6
A: Introns: 70/2; 160/3; 205/3; 249/3; 293/2; 350/1; 386/3; 452/3; 546/2; 669/3; 680/3

```

Query Match          19.08; Score 547.5; DB 2; Length 745;
Best Local Similarity 27.5%; Pred. No. 1.2e-34;
Matches 165; Conservative 105; Mismatches 243; Indels 87; Gaps 17;

QY      3 DYDEVTAFLGEGWGFQRLIFLFLSASIIIPNGFTGLGSVVELIATPEHRCRYP----- 53
      | : : : : | : : : : | : : : : | : : : : | : : : : |
Db     112 DFEGILNIIGCRWWQIWIYVIALQOIPIHMFNLNVVYMYDFEFOCMVPGFNNDTDTT 171
      | : : : : | : : : : | : : : : | : : : : | : : : : |
QY     54 -----DAANIISAWRNHVTPLRLRDGREVPHSC-----RRYRLATIANFSA 94
      | : : : : | : : : : | : : : : | : : : : | : : : : |
Db     172 ISSHGPLYLWGVDDIKNISVFEPNAN-----SDGAYQRDSCVFYERSEERYQLRRMPLET 226
      | : : : : | : : : : | : : : : | : : : : | : : : : |
QY     95 LGLPEGRDVLGLQLEQESCLDGEFFSDVYLSRTIVTEBNLVCEDDWKAPLITSLFF-VGV 153
      | : : : : | : : : : | : : : : | : : : : | : : : : |
Db     227 AMSEAWKDV-----APKKKC-QATHEKXDMVETITVDFNLVC-DSWFAKGHAHMFYSIGY 280
      | : : : : | : : : : | : : : : | : : : : | : : : : |
QY     154 LLGSFTISGQLSDRFGKKNVLFTVMGTGFSFLQI-----FSKNFEMFWVLVFLVGMG 206
      | : : : : | : : : : | : : : : | : : : : | : : : : |
Db     281 LLGCVLGGIASDKIGRKPTII-----GFGILSSMLGVLPENDYYPMELLIRLLSAIC 333
      | : : : : | : : : : | : : : : | : : : : | : : : : |
QY     207 QISNYVAAPVLCITTEILGKSVRIIFSTLGCVCIFAFGYMWVLPFAYFIRDRMILLVALTMP 266
      | : : : : | : : : : | : : : : | : : : : | : : : : |
Db     334 NEAADLAAYTLCEMITGTQYKRVAGSM-LQAPWALGYALLAIYATLTKSWKTIQVTAAGL 392
      | : : : : | : : : : | : : : : | : : : : | : : : : |
QY     267 GVLICVALWMPISPESRWLISQGFPEAEVITIRKAAKANGIVPSTIFD-----PS 316
      | : : : : | : : : : | : : : : | : : : : | : : : : |
Db     393 HFWSIIIFCIPESRWLVQNRVSAEEVIRKACR-----PPFPFNCTTSCGNLPS 447
      | : : : : | : : : : | : : : : | : : : : | : : : : |
QY     317 ELQDLASKKQOSH-----NILDLRTWNIRWVTIMSLMWTISVGFGLSL-----D 364
      | : : : : | : : : : | : : : : | : : : : | : : : : |
Db     448 DLELVSHRERKLNKNGKGTGFDLDLTMBKELRYTISVCIVFNATALVYVYGVNALSQDS 507
      | : : : : | : : : : | : : : : | : : : : | : : : : |
QY     365 TPN--LHGDIYFNCFLSAMVEVPAYVLAWLQLQYLPARYSMATALFLGGSVLLFMQLVP 421
      | : : : : | : : : : | : : : : | : : : : | : : : : |
Db     508 APQRTLTGTGFHLNNGIAGAEIPTLFACVMMQLGRKALMLTLTISGLFIIVAMLSMV 567
      | : : : : | : : : : | : : : : | : : : : | : : : : |
QY     422 PDLYIYATVLVMYKGFVTAAFSMWYVYVYTAELYPTVVRNNGVGSVSTASRLGSILSPYFV 481
      | : : : : | : : : : | : : : : | : : : : | : : : : |
Db     568 SGHYMLALAFMYFGKIAVGAFNLIYFTSELYPTVVRNTAVGVTSMWARFSGSLSSYIA 627
      | : : : : | : : : : | : : : : | : : : : | : : : : |
QY     482 YLGAYD-RELPTYLMGSLIITAILTLFLPESGTPLPDITDQMLR-VKGMKHKRTPSHT 539
      | : : : : | : : : : | : : : : | : : : : | : : : : |
Db     628 LUSNISLPVPMILFAVFLSFMAGLVFLPEYSEKLPETLDDAINFLPETKQFKT IEST 687

```

RESULT 8
T01019
transport protein homolog YUP8H12R.2 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 05-Feb-1999 #sequence_revision 05-Feb-1999 #text_change 22-Oct-1999
C:Accession: T01019
R:Theologis, A.; Vysotskaya, V.S.; Osborne, B.I.; Schwartz, J.R.; Federspiel, N.A.; K
Oefner, P.; Davis, R.W.
submitted to the EMBL Data Library, May 1998
A:Description: Arabidopsis thaliana chromosome 1 YAC YUP8H12R sequence.
A:Reference number: Z14227
A:Accession: T01019
A:Status: translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-527 <THE>
A:Cross-references: EMBL:AC002986; NID:g2494106; PID:g3152576; GSPDB:GN000059; ATSP:YU
A:Experimental source: cultivar Columbia
C:Genetics:
A:Gene: ATSP:YUP8H12R.2

RESULT 9
C96758
probable protein transporter Tl8K17.11 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: C96758
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.R.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719
A:Accession: C96758
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-539 <STO>
A:Cross-references: GB:AE005173; NID:g6598860; PIDN:AAF18714.1; GSPDB:GN00141
C:Genetics:
A:Gene: Tl8K17.11
A:Map position: 1

RESULT 10
B86299
hypothetical protein AAD34691.1 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: B86299
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, T.H.; Dewar,
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzia
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719
A:Accession: B86299
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-518 <STO>
A:Cross-references: GB:AE005172; NTD:94966360; PIDN:AAD34691.1; GSPDB:CN00141
C:Genetics:
A:Map position: 1

Query Match 16.3%; Score 470.5; DB 2; Length 518;
Best Local Similarity 28.2%; Pred. No. 7.4e-29;
Matches 136; Conservative 94; Mismatches 218; Indels 35; Gaps 12;

QY 55 AANLSSAW---RNHTVPLRLRGREVPHSCRRYRLATIANFSAIGLPEPGRDVLGQLEOE 111
DB 44 AALVSFAFEDAQQTITVFETDSQPMWHCONSDRVSCVNTSSNL-----89

QY 112 SCL---DGEFESQDYLSTIVTENLVCEDDWKAPLITSLFFVGVLLGSFISQQLSD-RF 167
DB 90 -CTLPNOTWSMDLPHVS-IIEKGLQACAGSFLKGFPPASSFFLGLGTLGSLALSTLADSSL 147

QY 168 GRKNVLFVTMGMTGCFSLQIFSKNFEMFVFLVFLVGMGOISNVAAFLVGTILKSKVR 227
DB 148 GRKNLLLSCLIMSLSMLTAFSTSIWYAFELFLNCGGRATGTCTALVSTELVKKWR 207

QY 228 IIFSTLGVCIYAF--GYMVLPLPAYIT--RDWRMLLVALTMPGVLCVALMWFIPESPR 282
DB 208 --GOVGAMGFCTGLFSLPLMLGYINEGNSRNLVYWTSTPTLIYCCLVRSFVRESPR 264

QY 283 WLISOGREFEAEVIRKAAKANGIVVPTIFDPSELQDLSKKQKQSHNLDLL--RTWNI 340
DB 265 WLIVKREAEVASILQSTA-SNAITMSETNLCFEVNDQSKNPVDYDALKILVRKSWSF 323

QY 341 RMYTMSIMLMTTSVGYGLSLDTPNLHGDIFVNCFLSAMVEYPAVYLAWLLOLPLRR 400
DB 324 RRL-LAAMVVGFGIGVYGYMPLALTNLNPLVGVVFNALSEFPAPLITFFFDIKNRR 382

QY 401 YSMATALLFGGSVLLFMQLVPPDLYLATVLMVKGKGVTAAPSMVYVTAELYPVVRN 460
DB 383 DALIGFTALSGLSALIAVLGQQLGSLQIVLELVSFSSACTAFNMTLIYTIEMFPTCVRN 442

QY 461 MGVGYSSTASRLGILSPFYVILGAYDRFLPYILMGSILITLITLFLPESFGTLPDT 520
DB 443 SAISMVRQALVFGVFSVPMVAAGRENQFWSYGLFGLIGLGVFLPETRGSLVCDT 502

QY 521 IDQ 523
DB 503 MDE 505

RESULT 11
T21682
hypothetical protein F32H5.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T21682
R:Steward, C.
submitted to the EMBL Data Library, November 1996
A:Reference number: 219458
A:Accession: T21682
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-528 <WIL>
A:Cross-references: EMBL:281524; PIDN:CAB04250.1; GSPDB:GN00023; CESP:F32H5.4
A:Experimental source: clone F32H5
C:Genetics:
A:Gene: CESP:F32H5.4
A:Map position: 5
A:Introns: 63/1; 108/3; 148/3; 249/2; 408/1; 472/3

Query Match 14.3%; Score 411; DB 2; Length 528;
Best Local Similarity 26.4%; Pred. No. 3.1e-24;
Matches 126; Conservative 91; Mismatches 169; Indels 92; Gaps 18;

QY 100 GRDVLGQLEQESCLDGEF-----SQDYLSITIVTENLVCEDDWKAPLISLF 149
DB 73 GPDVFI----KDIKCSQDFDHPNPNCTITSKYDFI--SINVDYGHFCGEGAWKTSISVQ 126

QY 150 FVGVLLGSFISQQLSDRFGRKNVLFVTMGMTGCFSLQIFSKNFEMFVFLVFLVGM--G 206
DB 127 MVGVLLGSVTSGAVADRYGRKLVLSVCFEMVSSLSILNTFAKDLIVFTIIRTILSIFKGG 186

QY 207 QISNVAAFLVGLGTEILKSKSVRIIFSTLGVGCIFYAFGVMVLPPLPAYFIRDW---RMLLYAL 263
DB 187 LLSTY---GVTKMEHPVPROHFWIATM---ISWAPYMILSFVAYLCHDWTYQYAIKAL 240

QY 264 TMPGVLCVALMWFIPESPRWLISOGREFEAEVIRKAAKANGIV-----VPSTIFDPS 316
DB 241 SMPGAV---VPLFKESPRWLIOAGRIEAEAROVULKHIMEVDGNTSDHSNNEIEMLOTEK 297

QY 317 ELQDLSKKQKQSHNLDLRTNIRMYTMSIMLM---TISVGYFGLSLDTPNLHGDIF 373
DB 298 KRQERIKRKNYDFRHLF--WNKYMASV-TMILWLMFSTSTNYGFVNEIKLSGSLY 354

QY 374 VNCFLSAMVEYPAVYLAWLLOLQYLPYRSMTALELGGSVLLFMQLVPPDLYLATVLM 433
DB 355 INALMGS-----LRWL-----NIFGLADLKFAKGRKHILSKLTTT 395

QY 434 VKRGV-----TAAFSMVYVYTA-----ELYPTVVRNMGVGS 466
DB 396 ICVFSIFVYVFEYEDYSLIIRVATLLASATASQVFTKSMVLMFEYPTVIRNSAVSEK 455

QY 467 STASRLGILSPFYVILGAYDRFLPYILMGSILITLITLFLPESFGTLPDITDQ 523
DB 456 SSASRIGTILGPQLFILCPY-KSLPYAILTGFCLEDAIAFQRLPETKGRPLPTMPE 512

RESULT 12
T25851
hypothetical protein T01B11.5 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T25851
R:Geisler, C.; Stellyes, L.
submitted to the EMBL Data Library, December 1996
A:Description: The sequence of C. elegans cosmid T01B11.
A:Reference number: Z20099
A:Accession: T25851
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-540 <GEI>
A:Cross-references: EMBL:U80931; PIDN:AAB38002.1; GSPDB:GN00022; CESP:T01B11.5
A:Experimental source: strain Bristol N2; clone T01B11
C:Genetics:
A:Gene: CESP:T01B11.5
A:Map position: 4
A:Introns: 85/1; 128/3; 209/2; 269/2; 352/2; 381/3; 489/3

Query Match 13.3%; Score 382.5; DB 2; Length 540;
Best Local Similarity 26.6%; Pred. No. 5.2e-22;
Matches 120; Conservative 78; Mismatches 178; Indels 75; Gaps 14;

QY 127 TIVTENLVCEDDWKAPLITSLFFVGVLLGSFISQQLSDRFGRKNVLFVTMGMTGFSFL 186
DB 124 SVQVEFNVICDARKVKNTITVTQFVLVGAATFGQVSDTFFGRKALLISTLGLNGLFNWI 183

QY 187 QIFSKNFEMFVFLVFLVGMGOISNVAAFLVGTILKSKSVRIIFSTLGVCIYAFGVMVL 246
DB 184 TAYSPDLFYMVWRTLAGVFTGGVTVVQVFMVENIPRKRDMWION---SIWSPNLIIF 240

QY 247 PLFAYFIRDWRMMLLVALTMPGVLCVALMWFIPESPRWLISOGREFEAE---VIRKAAKA 303
DB 241 PFVAVLQCDWRTMSVWIAAASIAITFLACLILESPRWLIQKGRIEEARKSLIRKTDK- 299

QY 304 NGIVVPSTIFD-----PSELQDLSKKQKQSHNLDLRTWNI--RMVTMSI 348
DB 300 -----LYDETEKOLDEVLHVEAEKHARSSKKAKKYTFIHLCTWKMIAQSFTTIG 351

QY 349 MLWMTISVGYFGLSLDTPNLHGDIFVNC-----FLSAMVEYPAVYLAWLLOLPLRR 400
DB 352 ITCTFTIV--YSLMNNKESGLYWNLAITMGASRIINIAVSADIYKLPWFGRKMI-NQ 408

QY 401 YSMATALLFGGSVLLFMQLVPPDLY--YLATVLMVKGKGVTAAPSMV-----YVYTA 451

Db 409 IAMICTLFA-----LLMIAVLYFYNGQLIAI-----GTVAASVCMCSOLFIAKYMVMVN 457
QY 452 ELYPTVVRNMGVGSSTASRLGSTLSPYFVYLGAIDRFLPYILMGSILTILT-AILTLELP 510
Db 458 ELYPTAVRNLAASVATMSRIGSFNSQLYLSDYAEWIPYAVLFSCQLYDFVLSLFLP 517
QY 511 ESFGTPLPDTIDQMLRVKMGKHKRTKTPSHTRM 541
Db 518 ETKGVIL-----ENHLPPKHKRI 535

RESULT 13
B96825
hypothetical protein T8K14.17 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719
A:Accession: B96825
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-515 <STO>
A:Cross-references: GB:AE005173; NID:g4835768; PIDN:AAD30235.1; GSPDB:GN00141
A:Gene: T8K14.17
A:Map position: 1

Query Match 13.2%; Score 380.5; DB 2; Length 515;
Best Local Similarity 26.7%; Pred. No. 7e-22;
Matches 118; Conservative 95; Mismatches 186; Indels 43; Gaps 15;

QY 113 CLD-----GHEFSQDYVLTIVTENLVCEDDWKAPLTLSLFEVGVLL 155
Db 72 CLDHTICNPATTDICKIPRSAMDGGFGKSVISEFDECCSSFLRSLPSSFTFYGSIV 131
QY 156 GSFISQSLD-RFRGNVLFVT---MGMTGFSFLQIFSKNFEMFVFLVGVGMQISNY 211
Db 132 GGVVLAIMPDSGLGRKQLLFSSFANSL-TGISI--FLSSNIWIYSFLKFGVGFARSQTG 188
QY 212 VAAVVLGTIELGKSVRIITSTIGVCIFYAGVGMVLPFLPAYFIR--DWRMLVALTMP-GV 268
Db 189 TVALVLISERISTKWR-PRATVPTLFLVGLMSLGLIAYLVRAHSMKVLVLTSPAGI 247
QY 269 LCVALWVFPSPRWLISOGREEAEVIRKAAKANGIVVPSTIFDPSELQDLSKKQ-Q 327
Db 248 HSIFYFFALESPRHLHLEGKKEAEVLKRISPNRNGYLESV---SRLRPKFTLEQTS 304
QY 328 SHNILD-LRTWNIRMTVMSLWMNTISVGVFGLSLDTPNLHGDFVNCFLSAMVEVPA 386
Db 305 SVSIKDLFIKNAFRVILVIMF-GLGMSYGVYPLAVRDIKVNIYSEALNANVELPT 363
QY 387 YVLAWLLQYLPYPRYSMAFALFGGS--VLLFMQLVPPDLY---YLATVLMVGVKFGVTA 441
Db 364 FVVTPILEQFSRRSVLVNCLIGGASGVLCFVM---SLYGRKIAFALELGSPFCARI 419
QY 442 AFSMWVYVYTAELPTVVRNMGVGSSTASRLGSLSPYFVYLGAIDRFLPYILMGSILIL 501
Db 420 GFNLAIYVELFPTCVRSATMMLRQALVVGACCPLIASIGRYIPSVSFAIFGMSGLGMFVLILP 494
QY 502 TAILTLFPESTGTPLPDIDQ 523

Db 480 LGLFALLPETKGLSLCDTME 501

RESULT 14
H86298
hypothetical protein AAD34689.1 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzia,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719
A:Accession: H86298
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-521 <STO>
A:Cross-references: GB:AE005172; NID:g4966358; PIDN:AAD34689.1; GSPDB:GN00141
C:Genetics:
A:Map position: 1

Query Match 12.8%; Score 368.5; DB 2; Length 521;
Best Local Similarity 25.4%; Pred. No. 6.1e-21;
Matches 110; Conservative 94; Mismatches 180; Indels 49; Gaps 14;

QY 117 WEFSQDYVLTIVTENLVCEDDWKAPLTLSLFEVGVLLGSFISGQL-SDRGRKN-VLF 174
Db 98 WEWDGSGQKSVISEFGLCESSLLRGMPSFAFYIGAIVGGFFLALIPDDSLGRKKLVLF 157
QY 175 VTGMQTFGSFLQIFSKNFEMFVFLVGVGMQISNVAAVFLGTIELG-----KSVRII 229
Db 158 STFAMSI-TSIVSFSTNVWIYTFLEKFIIGFSRSQTSVALVLSERVSTRWRPRATMIP 216
QY 230 FSTLGVCIIFYAGVGMVLPFLPAYFIR--WRMLVALTMPGV-LVALWVFPSPRWLS 286
Db 217 FT-----LVVLGMSLGLIAYLAQDSSWRYLYLTSVPAVFCIFYLFLFALESPLMHM 270
QY 287 QGRFEER-EVLIIRKAAKANGIVVPSTIFDPSELQDLSKKQKQSHNILD-LLRTWNIRMT 344
Db 271 QGDKKEAIDVLTMSPKKAYL--ESVSKLPLKQENFEQAPYISINDFFRKAFFRIL 328
QY 345 IMSLIMWMTISVGVFGLSLDTPNLHGDFVNCFLSAMVEVPAVYVLAWLLOYLPRYSMA 404
Db 329 VVLMIMF-GLGISYGVYPLAARDIDVNIYLSLSETLNALVELPTEVITPILLERENRSSLV 387
QY 405 TALFLGGS--VLLFMQLVPPDLYLATVLMVGVK-----FGVTAASFMYVYT 450
Db 388 VNTLGGASGVLCF-----VLSILGKTEIAFAFELGTFFCARIGFNLMVEM 434
QY 451 AELPTVVRNMGVGSSTASRLGSLSPYFVYLGAIDRFLPYILMGSILTILTALTILFLP 510
Db 435 VEMPTCVRSATMMLRQALVVGACCPLIASIGRYIPSVSFAIFGMSGLGMFVLILP 494
QY 511 ESFGTPLPDTIDQ 523
Db 495 ETKGLSLCDSMEE 507

RESULT 15
T23190
hypothetical protein ZK637.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 02-Jun-2000
C:Accession: T23190
R:Craxton, M.

submitted to the EMBL Data Library, April 1993
 A:Reference number: Z19704
 A:Accession: T23190
 A:Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: DNA
 A:Residues: 1-529 <WIL>
 A:Cross-references: EMBL:Z22175; PIDN:CAA80131.1; GSPDB:GN00021; CESP:2K637.1
 A:Experimental source: clone K01F9
 C:Genetics:
 A:Gene: CESP:ZK637.1
 A:Map position: 3
 A:Introns: 31/2; 93/3; 121/3; 150/3; 192/2; 255/3; 422/2; 472/3
 C:Superfamily: Caenorhabditis elegans glucose transport protein

	Query Match	12.3%	Score 355.5;	DB 2;	Length 529;
	Best Local Similarity	26.5%	Pred. No. 6.3e-20;		
	Matches 140;	Conservative 97;	Mismatches 186;	Indels 105;	Gaps 24;
QY	52	VPDAANLSSANRNHTVPLRLRDGREVPHSCRRYRLATIANFSAIGLEP--GRDVD-----	104		
Db	10	VLEASNLTEAVYDITAKOLKEIRHVGDDF-AVRYSNLDDRTLEG-EPTDQRSPDSEKTF	67		
QY	105	-----LCOLEQE-SCLDGWFEFSQD-----VYLSTI-----VTEWNLVCEDDWKAPLT	145		
Db	68	TVDEAVEALGFRQLKLSILTGMAWADAMENMLLSISPALACEWGI---SSVQQALV	124		
QY	146	ISLFPVGLGSGFSGQLSDRFG-RKNVLFVT-----MGMTGFS---FLQIFSKNFEMF	196		
Db	125	TCVFSGMMLSTFWGKICDRGRKGLTETLVACIMGVISGMSPHFVLLFFRGLTGF	184		
QY	197	VVLFVLGVMGOISHYVAAFV--LGTEILGKSVRIETSLGVCIFYAFGYMVLPLFAYFIR	254		
Db	185	-----GIGGVPQSVTIYAEFLTAQRAKCVLIES-----FWAIGAVFEALLAYFVM	231		
QY	255	D---WRMLLVALTNP-GVLCVALMWPIPESPRWLISQGRFEAEVIRKAAKANGIVWPS	310		
Db	232	ESFGWRALMFLSSGULGIFAVAFW-LPESARFDMASGHPERALETLQAAARNRVQLPT	290		
QY	311	TIFDPSELQDLSSKKQOSHNLIDL--LRTWNIRMTVIMSIMLWMTISVGYFGLSLDTPNL	368		
Db	291	G-----RLVSSIKAGSESGDIANLLSPDLRKTILLWCITAITAFSYGVMVLFETVL	343		
QY	369	-----HGDIFVN-----CFLSAMVEVPAYVLAWLLQYLPFRYSMA	404		
Db	344	FQSHDECHGGLFSRGTMQEVQCPLTRSDYFDLLSTTLAEPPGLIITVLIIEWFGRKKTMA	403		
QY	405	T--ALFLGSGVLLFMQLVPPDLYLATVLMVGKFGVTAAFSMVYVYTAELYPTVVRNMG	462		
Db	404	LEYAVFAITFELLYFCLD----RFTVVVLIFVARAFISGAFQCAVYVTPPEVPTTLRANG	459		
QY	463	GVGVSTASRLGSLISPYFVYLGAYDRFLPYILMGSLTILTAITLFLP	510		
Db	460	LGTCSAMARIGAIVA-----SEKSLPIGIYGTAAITAILGLIASLSLP	501		

Search completed: August 16, 2001, 13:59:29
 Job time: 156 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 16, 2001, 14:04:00 ; Search time 18.15 Seconds
(without alignments)
1051.257 Million cell updates/sec

Title: US-09-521-195-3
Perfect score: 2883
Sequence: 1 MRDYDEVTAFLGNGFFQRL.....HTRMLKQGERPTILKSTAF 557

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %		Length	DB ID	Description
		Match	Match			
1	2883	100.0	557	1	OCN2_HUMAN	O76082 homo sapien
2	2513	87.2	557	1	OCN2_MOUSE	Q920e8 mus musculu
3	2509	87.0	557	1	OCN2_RAT	O70594 rattus norv
4	387.5	13.4	751	1	YX5_CAEEL	P46501 caenorhabdi
5	357	12.4	1222	1	YMP3_CAEEL	Q10947 caenorhabdi
6	355.5	12.3	529	1	Y0U1_CAEEL	P30638 caenorhabdi
7	308	10.7	445	1	YGC5_ECOLI	Q46909 escherichia
8	291.5	10.1	435	1	Y113_CAEEL	Q10917 caenorhabdi
9	287	10.0	400	1	YCE1_BACSU	O34691 bacillus su
10	267	9.3	443	1	YAAU_ECOLI	P31679 escherichia
11	265	9.2	459	1	YDUK_ECOLI	P76230 escherichia
12	262.5	9.1	742	1	SVY2_RAT	Q02563 rattus norv
13	256	8.9	482	1	YF1G_BACSU	P54723 bacillus su
14	255.5	8.9	452	1	YDJE_ECOLI	P38055 escherichia
15	252.5	8.8	457	1	PCAK_ACICA	Q43975 acinetobact
16	250.5	8.7	472	1	ARAE_KLEOX	P45598 klebsiella
17	243.5	8.4	551	1	HGT1_KLEULA	P49374 kluyveromyc
18	241.5	8.4	472	1	ARAE_ECOLI	P09830 escherichia
19	240.5	8.3	495	1	GTR3_CANFA	P47842 canis fami
20	239.5	8.3	451	1	YVAT_BACSU	P37514 bacillus su
21	239	8.3	461	1	CSPC_BACSU	P46333 bacillus su
22	237.5	8.2	496	1	GTR3_HUMAN	P11169 homo sapien
23	234	8.1	494	1	GTR3_SHEEP	P47843 ovis aries
24	227	7.9	490	1	GTR1_CHICK	P46896 gallus gall
25	226.5	7.9	464	1	GALP_ECOLI	P37021 escherichia
26	224.5	7.8	413	1	MUCK_ACICA	P94131 acinetobact
27	223	7.7	451	1	GTR1_PIG	P20303 sus scrofa
28	222	7.7	763	1	RGY2_YEAST	Q12300 saccharomyc
29	221	7.7	492	1	GTR1_RAT	P11167 rattus norv
30	220	7.6	492	1	GTR1_BOVIN	P27674 bos taurus
31	219.5	7.6	491	1	XYLE_ECOLI	P09098 escherichia
32	219	7.6	493	1	GTR3_MOUSE	P32037 mus musculu
33	217	7.5	496	1	GTR3_CHICK	P28568 gallus gall

34	216	7.5	492	1	GTR1_HUMAN	P11166 homo sapien
35	216	7.5	492	1	GTR1_MOUSE	P17809 mus musculu
36	215	7.5	492	1	GTR1_RABIT	P13355 oryctolagus
37	213.5	7.4	486	1	YGR4_YEAST	P53142 saccharomyc
38	213	7.4	448	1	PCAK_PSEPU	Q51955 pseudomonas
39	213	7.4	818	1	SNF3_YEAST	P10870 saccharomyc
40	208	7.2	493	1	GTR3_RAT	Q07647 rattus norv
41	207.5	7.2	567	1	HXT9_YEAST	P40885 saccharomyc
42	206.5	7.2	567	1	HXTA_YEAST	P54862 saccharomyc
43	204	7.1	547	1	GTR1_DEIDO	Q01440 leishmania
44	201	7.0	592	1	HXT5_YEAST	P36695 saccharomyc
45	200.5	7.0	522	1	STP1_ARATH	P23586 arabidopsis

ALIGNMENTS

RESULT	1
OCN2_HUMAN	
ID	OCN2_HUMAN
AC	O76082; STANDARD; PRT; 557 AA.
DT	01-OCT-2000 (Rel. 40, Created)
DT	01-OCT-2000 (Rel. 40, Last annotation update)
DT	01-OCT-2000 (Rel. 40, Last annotation update)
DE	ORGANIC CATION/CARNITINE TRANSPORTER 2 (SOLUTE CARRIER FAMILY 22, MEMBER 5) (HIGH-AFFINITY SODIUM-DEPENDENT CARNITINE COTRANSPORTER).
DE	SIC22A5 OR OCTN2.
GN	Homo sapiens (human).
OS	Homo sapiens (human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX	NCBI_TaxID=9606;
RN	[1]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=98289574; PubMed=9618255;
RA	Wu X., Prasad P.D., Leibach F.H., Ganapathy V.;
RT	*CDNA sequence, transport function, and genomic organization of human OCTN2, a new member of the organic cation transporter family.*;
RL	Biochem. Biophys. Res. Commun. 246:589-595(1998).
RN	[2]
RP	SEQUENCE FROM N.A.
RX	TISSUE=Kidney;
RC	MEDLINE=98352077; PubMed=9685390;
RA	Tamai I., Ohashi R., Nezu J.-I., Yabuuchi H., Oku A., Shimane M., Sai Y., Tsuji A.;
RT	*Molecular and functional identification of sodium ion-dependent, high affinity human carnitine transporter OCTN2.*;
RL	J. Biol. Chem. 273:20378-20382(1998).
RN	[3]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=99113835; PubMed=9916797;
RA	Nezu J., Tamai I., Oku A., Ohashi R., Yabuuchi H., Hashimoto N., Nikaudo H., Sai Y., Koizumi A., Shoji Y., Takada G., Matsuishi T., Yashino M., Kato H., Ohura T., Tsujimoto G., Hayakawa J., Shimane M., Tsuji A.;
RT	*Primary systemic carnitine deficiency is caused by mutations in a gene encoding sodium ion-dependent carnitine transporter.*;
RL	Nat. Genet. 21:91-94(1999).
RN	[4]
RP	CHARACTERIZATION.
RA	MEDLINE=99384224; PubMed=10454528;
RX	Wu X., Huang W., Prasad P.D., Seth P., Rajan D.P., Leibach F.H., Chen J., Conway S.J., Ganapathy V.;
RT	*Functional characteristics and tissue distribution pattern of organic cation transporter 2 (OCTN2), an organic cation/carnitine transporter.*;
RL	J. Pharmacol. Exp. Ther. 290:1482-1492(1999).
RN	[5]
RP	VARIANT CDSP GLN-169.
RX	MEDLINE=99355597; PubMed=10425211;
RA	Burwinkel B., Kreuder J., Schweitzer S., Vorgerd M., Gempel K., Gerbitz K.-D., Killmann M.W.;
RT	*Carnitine transporter OCTN2 mutations in systemic primary carnitine deficiency: a novel Arg169Gln mutation and a recurrent Arg282ter

RT mutation associated with an unconventional splicing abnormality.";
 RL Biochem. Biophys. Res. Commun. 261:484-487(1999).
 RN [6]
 RP VARIANT CDSP CYS-211.
 RX MEDLINE=99408248; PubMed=10480371;
 RA Vaz F.N., Scholte H.R., Ruiter J., Hussaarts-Odijk L.M.,
 RA Rodrigues Pereira R., Schweitzer S., de Klerk J.B.C., Waterham H.R.,
 RA Wanders R.J.A.;
 RT "Identification of two novel mutations in OCTN2 of three patients with
 RT systemic carnitine deficiency.";
 RL Hum. Genet. 105:157-161(1999).
 RN [7]
 RP VARIANT CDSP LEU-478.
 RX MEDLINE=99172075; PubMed=10072434;
 RA Tang N.L., Ganapathy V., Wu X., Hui J., Seth P., Yuen P.M.,
 RA Wanders R.J., Fok T.F., Hjelm N.M.;
 RT "Mutations of OCTN2, an organic cation/carnitine transporter, lead to
 RT deficient cellular carnitine uptake in primary carnitine deficiency.";
 RL Hum. Mol. Genet. 8:655-660(1999).
 RN [8]
 RP CHARACTERIZATION OF VARIANT CDSP LEU-478, AND MUTAGENESIS.
 RX MEDLINE=20026865; PubMed=10559218;
 RA Seth P., Wu X., Huang W., Leibach F.H., Ganapathy V.;
 RT "Mutations in novel organic cation transporter (OCTN2), an organic
 RT cation/carnitine transporter, with differential effects on the
 RT organic cation transport function and the carnitine transport
 RT function.";
 RL J. Biol. Chem. 274:33388-33392(1999).
 RN [9]
 RP VARIANTS CDSP ARG-383 AND PHE-446.
 RX MEDLINE=20081068; PubMed=10612840;
 RA Mayatepek E., Nezu J., Tamai I., Oku A., Katsura M., Shimane M.,
 RA Tsuji A.;
 RT "Two novel missense mutations of the OCTN2 gene (W283R and V446F) in a
 RT patient with primary systemic carnitine deficiency.";
 RL Hum. Mutat. 15:118-118(2000).
 RN [10]
 RP VARIANT CDSP LYS-452.
 RX MEDLINE=20145665; PubMed=10679939;
 RA Wang Y., Kelly M.A., Cowan T.M., Longo N.;
 RT "A missense mutation in the OCTN2 gene associated with residual
 RT carnitine transport activity.";
 RL Hum. Mutat. 15:238-245(2000).
 CC -!- FUNCTION: SODIUM-ION DEPENDENT, HIGH AFFINITY CARNITINE
 CC TRANSPORTER. ALSO TRANSPORTS ORGANIC CATIONS WITHOUT THE
 CC INVOLVEMENT OF SODIUM. INVOLVED IN THE ACTIVE CELLULAR UPTAKE OF
 CC CARNITINE.
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -!- TISSUE SPECIFICITY: STRONGLY EXPRESSED IN KIDNEY, SKELETAL MUSCLE,
 CC HEART AND PLACENTA.
 CC -!- DISEASE: DEFECTS IN SLC22A5 ARE THE CAUSE OF SYSTEMIC PRIMARY
 CC CARNITINE DEFICIENCY (CDSP). CDSP IS AN AUTOSOMAL RECESSIVE
 CC DISORDER OF FATTY ACID OXIDATION CAUSED BY DEFECTIVE CARNITINE
 CC TRANSPORT. PRESENT EARLY IN LIFE WITH HYPOKETOTIC HYPOGLYCEMIA AND
 CC ACUTE METABOLIC DECOMPENSATION, OR LATER IN LIFE WITH SKELETAL
 CC MYOPATHY OR CARDIOMYOPATHY.
 CC -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY. ORGANIC
 CC CATION SUBFAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF057164; AAC24828.1; -;
 DR EMBL; AB015050; BAA29023.1; -;
 DR EMBL; AB016625; BAA36712.1; -;
 DR MIM; 603377; -;
 DR MIM; 212140; -;
 DR InterPro; IPR001066; -;

DR InterPro; IPR001687; -;
 DR Pfam; PF00083; sugar_tr; 1.
 DR PROSITE; PS00216; SUGAR_TRANSPORT_1; 1. disease mutation.
 KW transport; transmembrane; Glycoprotein; POTENTIAL.
 FT TRANSSEM 21 41
 FT TRANSSEM 143 163
 FT TRANSSEM 173 193
 FT TRANSSEM 196 216
 FT TRANSSEM 233 253
 FT TRANSSEM 258 278
 FT TRANSSEM 343 363
 FT TRANSSEM 372 392
 FT TRANSSEM 414 434
 FT TRANSSEM 437 457
 FT TRANSSEM 489 509
 FT CARBOHYD 57 57
 FT CARBOHYD 64 64
 FT CARBOHYD 91 91
 FT VARIANT 169 169
 FT VARIANT 211 211
 FT VARIANT 283 283
 FT VARIANT 446 446
 FT VARIANT 452 452
 FT VARIANT 478 478
 FT MUTAGEN 352 352
 FT SEQUENCE 557 AA; 62751 MW; 928B1F6EFF63C48D CRC64;
 SQ
 Query Match 100.0%; Score 2883; DB 1; Length 557;
 Best Local Similarity 100.0%; Pred. No. 3.1e-178;
 Matches 557; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MRDDEVTAFLGEGPQRLIFLLSASIPNGFTGLSSVFLIATPEHRCRVDAANLS 60
 DB 1 MRDDEVTAFLGEGPQRLIFLLSASIPNGFTGLSSVFLIATPEHRCRVDAANLS 60
 QY 61 AWRNHTVPLRLDGRVPHSCRRYRLATIANFSALEPGRDVLQLEQESCLDGWEFS 120
 DB 61 AWRNHTVPLRLDGRVPHSCRRYRLATIANFSALEPGRDVLQLEQESCLDGWEFS 120
 QY 121 QDVYLSITVTENLVCEDDWKAPLITISLFFVGVLLGSFISGQLSDRGRKNVLFVTMGQ 180
 DB 121 QDVYLSITVTENLVCEDDWKAPLITISLFFVGVLLGSFISGQLSDRGRKNVLFVTMGQ 180
 QY 181 TGFSPLOIFSKNFEMFVFLVGVGMGOISNVAAFLVGLTGILGKSVRIIFSTLGVCFYFA 240
 DB 181 TGFSPLOIFSKNFEMFVFLVGVGMGOISNVAAFLVGLTGILGKSVRIIFSTLGVCFYFA 240
 QY 241 FGYNVPLFAYFIRDWRMLVALTMPGVLCVALMWFIPESPRWLISQGRFEEAEVIRKA 300
 DB 241 FGYNVPLFAYFIRDWRMLVALTMPGVLCVALMWFIPESPRWLISQGRFEEAEVIRKA 300
 QY 301 AKANGIVVPSTIFDSELDLSSKKQSHNLLDLRTWNIRMTVIMSIMLWMTISVGYFG 360
 DB 301 AKANGIVVPSTIFDSELDLSSKKQSHNLLDLRTWNIRMTVIMSIMLWMTISVGYFG 360
 QY 361 LSLDTPNLHGDIFFVNCFLSAMVEVPAYVLAWLQLQYLPRIYSMATALFLGGSVLLPMQLV 420
 DB 361 LSLDTPNLHGDIFFVNCFLSAMVEVPAYVLAWLQLQYLPRIYSMATALFLGGSVLLPMQLV 420
 QY 421 PPDLYLATVLMVGMKFGVTAAFSMVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVY 480
 DB 421 PPDLYLATVLMVGMKFGVTAAFSMVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVY 480

Qy 481 VYLGYDRFLPYILMGSLTILTAILTLFLPESFGTPLPDTIDQMLRVKGMKHKRTPSHTR 540
Db 481 VYLGYDRFLPYILMGSLTILTAILTLFLPESFGTPLPDTIDQMLRVKGMKHKRTPSHTR 540
Qy 541 MKDQGERPTILKSTAF 557
Db 541 MKDQGERPTILKSTAF 557
RESULT 2
OCN2_MOUSE STANDARD; PRT; 557 AA.
AC Q920E8;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE ORGANIC CATION/CARNITINE TRANSPORTER 2 (SOLUTE CARRIER FAMILY 22,
DE MEMBER 5) (HIGH-AFFINITY SODIUM-DEPENDENT CARNITINE COTRANSPORTER).
GN SLC22A5 OR OCTN2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Kidney;
RX MEDLINE=99113835; PubMed=9916797;
RA Nezu J., Tamai I., Oku A., Ohashi R., Yabuuchi H., Hashimoto N.,
RA Nikaide H., Sai Y., Koizumi A., Shoji Y., Takada G., Matsushita T.,
RA Yashino M., Kato H., Ohura T., Tsujimoto G., Hayakawa J., Shimane M.,
RA Tsuji A.;
RT "Primary systemic carnitine deficiency is caused by mutations in a
RT gene encoding sodium ion-dependent carnitine transporter.";
RL Nat. Genet. 21:91-94(1999).
RN [2]
RP SEQUENCE FROM N.A., AND VARIANT JVS ARG-352.
RC STRAIN=C3H;
RX MEDLINE=99057546; PubMed=9837751;
RA Lu K., Nishimori H., Nakamura Y., Shima K., Kuwajima M.;
RT "A missense mutation of mouse OCTN2, a sodium-dependent carnitine
RT cotransporter, in the juvenile visceral steatosis mouse.";
RL Biochem. Biophys. Res. Commun. 252:590-594(1998).
RN [3]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RX MEDLINE=99384224; PubMed=10454528;
RA Wu X., Huang W., Prasad P.D., Seth P., Rajan D.P., Leibach F.H.,
RA Chen J., Conway S.J., Ganapathy V.;
RT "Functional characteristics and tissue distribution pattern of organic
RT cation transporter 2 (OCTN2), an organic cation/carnitine
RT transporter.";
RL J. Pharmacol. Exp. Ther. 290:1482-1492(1999).
CC -1- FUNCTION: SODIUM-ION DEPENDENT, HIGH AFFINITY CARNITINE
CC TRANSPORTER. ALSO TRANSPORTS ORGANIC CATIONS WITHOUT THE
CC INVOLVEMENT OF SODIUM. INVOLVED IN THE ACTIVE CELLULAR UPTAKE OF
CC CARNITINE.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- DISEASE: DEFECTS IN SLC22A5 ARE THE CAUSE OF JUVENILE VISCERAL
CC STEATOSIS (JVS).
CC -1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY. ORGANIC
CC CATION SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB015800; BAA36590.1; -;
DR EMBL; AF111425; AAC99787.1; -;
DR EMBL; AF110417; AAD54060.1; -;
DR InterPro; IPR001066; -;

Pfam: PF00083; sugar_tr; 1.
DR PROSITE; PS00216; SUGAR_TRANSPORT_1; 1.
KW Transport; Transmembrane; Glycoprotein; Disease mutation.
FT TRANSMEM 21 41
FT TRANSMEM 143 163
FT TRANSMEM 173 193
FT TRANSMEM 198 218
FT TRANSMEM 233 253
FT TRANSMEM 258 278
FT TRANSMEM 342 362
FT TRANSMEM 374 394
FT TRANSMEM 407 427
FT TRANSMEM 431 451
FT TRANSMEM 489 509
FT CARBOHYD 57 57
FT CARBOHYD 64 64
FT CARBOHYD 91 91
FT CARBOHYD 322 322
FT VARIANT 352 352
SQ SEQUENCE 557 AA; 62779 MW; 6093F0EE9612B204 CRC64;
Query Match 87.2%; Score 2513; DB 1; Length 557;
Best Local Similarity 85.5%; Pred No. 1.7e-154;
Matches 476; Conservative 39; Mismatches 42; Indels 0; Gaps 0;
Qy 1 MRDYDEVTAFLGEGWPGFQRLIFLLSASIIIPNGFTGLSSVFLIATPEHRCRVPDANLSS 60
Db 1 MRDYDEVTAFLGEGWPGFQRLIFLLSASIIIPNGFTGLSSVFLIATPEHRCRVPDANLSS 60
Qy 61 AWRNHTVPLRLDGRVPHSCRRYRLATIANFSALEGRVDVLQLEQESCLDGEWERS 120
Db 61 AWRNHSIPLTKDGRVQPKRRYRLATIANFSELGLEPGRVDVLQLEQESCLDGEWYD 120
Qy 121 ODVYLSITVTEWNLVCEDDMKAPLTISLFEVGLVLSGSLSDRFRGNKLVLFVTGMQ 180
Db 121 KDVFSLITVTEWDLVCKDDMKAPLTISLFEVGLVLSGSLSDRFRGNKLVLFVTGMQ 180
Qy 181 TGSFLOIFSKNFEMFVFLVGVGMGQISNYAAFLVGLTEILGSKSVRIIFSTLGVCFYA 240
Db 181 TGSFLOVFSVNFEMFVFLVGVGMGQISNYAAFLVGLTEILSKSRIIFATLGVCFYA 240
Qy 241 FGYMVLPLFAYFIRDRMMLVALTMPGVLCVALWVFIPESPRWLIISQGRFEEAEVIRKA 300
Db 241 FGYMVLPLFAYFIRDRMMLVALTMPGVLCVALWVFIPESPRWLIISQGRFEEAEVIRKA 300
Qy 301 AKANGIVVPSTIFDPSQLSSKQKSHNLDLRTNIRMTVIMSLMWTISVGYFG 360
Db 301 AKINGIVAPSTIFDPSQLSSKQKSHNLDLRTNIRMTVIMSLMWTISVGYFG 360
Qy 361 LSLDTPNLHGDIYVNCFLSAVVEVPAYVLAWLLQYLPRYSMATALFLGSGVLLFMQV 420
Db 361 LSLDTPNLHGDIYVNCFLSAVVEVPAYVLAWLLQYLPRYSMATALFLGSGVLLFMQV 420
Qy 421 PPDLYLATVLMVGKFGVTAFAFVSVVYVYAEVPTVVRNMGVGVSTASRLSILSPYF 480
Db 421 PSELFYLTALVMVGKFGVTAFAFVSVVYVYAEVPTVVRNMGVGVSTASRLSILSPYF 480
Qy 481 VYLGYDRFLPYILMGSLTILTAILTLFLPESFGTPLPDTIDQMLRVKGMKHKRTPSHTR 540
Db 481 VYLGYDRFLPYILMGSLTILTAILTLFLPESFGTPLPDTIDQMLRVKGMKHKRTPSHTR 540
Qy 541 MKDQGERPTILKSTAF 557
Db 541 MKDQGERPTILKSTAF 557
RESULT 3
OCN2_MOUSE STANDARD; PRT; 557 AA.
AC Q920E8;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)

01-OCT-2000 (Rel. 40, Last annotation update)
 ORGANIC CATION/CARNITINE TRANSPORTER 2 (SOLUTE CARRIER FAMILY 22, MEMBER 5) (HIGH-AFFINITY SODIUM-DEPENDENT CARNITINE COTRANSPORTER) (UST2R) (CTI).
 GN SLC22A5 OR OCTN2.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus. NCBI_TaxID=10116;
 [1]
 RN SEQUENCE FROM N.A.
 RP TISSUE-Kidney;
 RX MEDLINE=98200080; PubMed=9541011;
 RA Schoenig E., Spitzberger F., Engelhardt M., Martel F., Oerding N., Gruendemann D.;
 RA "Molecular cloning and characterization of two novel transport proteins from rat kidney.";
 RT FEBS Lett. 425:79-86(1998).
 [2]
 RN SEQUENCE FROM N.A.
 RP STRAIN=SPRAGUE-DAWLEY; TISSUE=Intestine;
 RX MEDLINE=99011422; PubMed=9792817;
 RA Sekine T., Kusunahara H., Utsunomiya-Tate N., Tsuda M., Sugiyama Y., Kanai Y., Endou H.;
 RA "Molecular cloning and characterization of high-affinity carnitine transporter from rat intestine.";
 RT Biochem. Biophys. Res. Commun. 251:586-591(1998).
 [3]
 RN SEQUENCE FROM N.A.; AND CHARACTERIZATION.
 RX MEDLINE=99384224; PubMed=1045428;
 RA Wu X., Huang W., Prasad P.D., Seth P., Rajan D.P., Leibach F.H., Chen J., Conway S.J., Ganapathy V.;
 RA "Functional characteristics and tissue distribution pattern of organic cation transporter 2 (OCTN2), an organic cation/carnitine transporter.";
 RT J. Pharmacol. Exp. Ther. 290:1482-1492(1999).
 CC -1- FUNCTION: SODIUM-ION DEPENDENT, HIGH AFFINITY CARNITINE TRANSPORTER. ALSO TRANSPORTS ORGANIC CATIONS WITHOUT THE INVOLVEMENT OF SODIUM. INVOLVED IN THE ACTIVE CELLULAR UPTAKE OF CARNITINE.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN THE PROXIMAL AND DISTAL TUBULES AND IN THE GLOMERULI IN THE KIDNEY, IN THE MYOCARDIUM, VALVES, AND ARTERIOLES IN THE HEART, IN THE LABYRINTHINE LAYER OF THE PLACENTA, AND IN THE CORTEX, HIPPOCAMPUS, AND CEREBELLUM IN THE BRAIN.
 CC -1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY. ORGANIC CATION SUBFAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AJ001933; CAA05106.1; -
 CC EMBL; AB017260; BAA34399.1; -
 CC EMBL; AF110416; AAD54059.1; -
 CC InterPro; IPR001066; -
 CC Pfam; PF00083; sugar_tr.1;
 CC PROSITE; PS00216; SUGAR_TRANSPORT_1; 1.
 KW Transport; Transmembrane; Glycoprotein.
 FT TRANSMEM 21 41 POTENTIAL.
 FT TRANSMEM 143 163 POTENTIAL.
 FT TRANSMEM 173 193 POTENTIAL.
 FT TRANSMEM 198 218 POTENTIAL.
 FT TRANSMEM 233 253 POTENTIAL.
 FT TRANSMEM 258 278 POTENTIAL.
 FT TRANSMEM 342 362 POTENTIAL.
 FT TRANSMEM 374 394 POTENTIAL.
 FT TRANSMEM 407 427 POTENTIAL.

FT TRANSMEM 431 451 POTENTIAL.
 FT TRANSMEM 489 509 POTENTIAL.
 FT CARBOHYD 57 57 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 64 64 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 91 91 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 275 275 W -> G (IN REF. 2).
 SQ SEQUENCE 557 AA; 62567 MW; 172472E7B0A5F043 CRC64;
 Query Match 87.0%; Score 2509; DB 1; Length 557;
 Best Local Similarity 85.5%; Pred. No. 3.le-154;
 Matches 476; Conservative 42; Mismatches 39; Indels 0; Gaps 0;
 QY 1 MRDYDEVTAFLEGMGPPQRLIFELLSASIIIPNGFTGLSSVFLLIATPEHRCRVPEDAANLS 60
 DB 1 MRDYDEVTAFLEGMGPPQRLIFELLSASIIIPNGFTGLSSVFLLIATPEHRCRVPEHVTNLS 60
 QY 61 AWRNHTVPLRLDRGPHSCRRYRLATIANFSALEPRGVDVLGQLESCLDGWFEFS 120
 DB 61 AWRNHSIPLETGDRQVPOSCRRYRLATIANFSALEPRGVDVLGQLESCLDGWFEYN 120
 QY 121 ODVYLSITVTENLVCEDDKAPLTISLEFVGLGSLFSGQLSDRFGKKNVLFVTMGQ 180
 DB 121 KDVFSLTIVTENDLVCKDKKAPLTISLEFVGLGSLFSGQLSDRFGKKNVLFVTMGQ 180
 QY 181 TGFSEFLQIFSKNEMFVWLFVLVGMGOISNYVAAFVLGTTEILGKSVRIIFSTLGVCI 240
 DB 181 TGFSEFLQIFSVNEMFVWLFVLVGMGOISNYVAAFVLGTTEILGKSVRIIFATLGVCI 240
 QY 241 FGMVLPFLFAFTRDWRMLLVALTMPGVLCVALWPFIPSPRWLISQGRFEAEVIRKA 300
 DB 241 FGMVLPFLFAFTRDWRMLLVALTMPGVLCVALWPFIPSPRWLISQGRFEAEVIRKA 300
 QY 301 AKANGIVPSTIEDPSELQDLSSKKOOSHNDLLRTNIRMTVMTSMLMTISVGYFG 360
 DB 301 AKNGIVAPSTIEDPSELQDLSSKKOOSHNDLLRTNIRMTVMTSMLMTISVGYFG 360
 QY 361 LSLDTPNLHGDIIVNCFLSAMVEVPAYVLAWLLQYLPYRYRYMATATLGGSVLLFMOLV 420
 DB 361 LSLDTPNLHGDIIVNCFLSAMVEVPAYVLAWLLQYLPYRYRYMATATLGGSVLLFIOLV 420
 QY 421 PPDLVLYATVLMVVGKFGVTAFAFVSVYVTAELYPTVVRNMGVSVSTASRLGSLSPYF 480
 DB 421 PSELYLSTALVWVGFGITISAYSVMVYVTAELYPTVVRNMGVSVSTASRLGSLSPYF 480
 QY 481 VYLGAYDRFLPYILMGSLTILTAITLFLPESFGTLPDPTIDQMLRVKMGKHKRTPSHR 540
 DB 481 VYLGAYDRFLPYILMGSLTILTAITLFLPESFGTLPDPTIDQMLRVKMGKHKRTPSHR 540
 QY 541 MLKDGQRPILKSTAF 557
 DB 541 TQKDGGSPTVLKSTAF 557
 RESULT 4
 YLX5_CAEEL STANDARD; PRT; 751 AA.
 ID YLX5_CAEEL
 AC P46501;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE HYPOTHETICAL 84.8 KDA PROTEIN F23F12.5 IN CHROMOSOME III.
 GN F23F12.5
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RA Du Z.;
 RL Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).

DR	EMBL; UI2965; AAA20607.1; -. WormPep; F23F12.5; CE01252.
DR	InterPro; IPR001066; -. Pfam; PF00083; sugar_tr; 1.
KW	Hypothetical protein; Transmembrane.
FT	TRANSMEM 24 44 POTENTIAL.
FT	TRANSMEM 57 77 POTENTIAL.
FT	TRANSMEM 102 122 POTENTIAL.
FT	TRANSMEM 137 157 POTENTIAL.
FT	TRANSMEM 200 220 POTENTIAL.
FT	TRANSMEM 235 255 POTENTIAL.
FT	TRANSMEM 320 340 POTENTIAL.
FT	TRANSMEM 348 368 POTENTIAL.
FT	TRANSMEM 379 399 POTENTIAL.
FT	TRANSMEM 410 430 POTENTIAL.
FT	TRANSMEM 432 452 POTENTIAL.
FT	TRANSMEM 515 535 POTENTIAL.
FT	TRANSMEM 547 567 POTENTIAL.
FT	TRANSMEM 583 603 POTENTIAL.
FT	TRANSMEM 614 634 POTENTIAL.
FT	TRANSMEM 678 698 POTENTIAL.
SQ	SEQUENCE 751 AA; 848332 MW; A6C4F435A40295BFC CRC64;

Db 364 QTLALIGTVFL-----DSVEFKLIVMLVAKVMATIIYSVHPDWATEQPTSVRSICF 416

QY 464 GVSSTASRLGSLSPYVYVGLAYDFLPYILMGSLTILTAITLFLPESFGTLPDITDQ 523

Db 417 SLMNIPQSMGIIMSPVVKHIVSNPWIPVVLVFSISATLAFMLHETKNKLLPTDIES 476

QY 524 MLRVGKMKHKRTPSHRML 542

Db 477 L-----SYPSETNDL 486

RESULT 6

YOUNL_CAEEL STANDARD; PRT; 529 AA.

AC P30638; Q21101;

DT 01-APR-1993 (Rel. 25, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 30-MAY-2000 (Rel. 39, Last annotation update)

DE HYPOTHETICAL 58.3 KDA PROTEIN ZK637.1 IN CHROMOSOME III.

GN ZK637.1.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

OC Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI_TaxID=6239;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2;

RX MEDLINE=92168156; PubMed=1538779;

RA Sulston J., Du Z., Thomas K., Willier L., Staden R.,

RA Halloran N., Green P., Thierry-Mieg J., Qiu L., Dear S., Coulson A.,

RA Craxton M., Durbin R.K., Berks M., Metzstein M., Hawkins T.,

RA Ainscough R., Waterston R.;

RT "The C. elegans genome sequencing project: a beginning.";

RL Nature 356:37-41(1992).

RN [2]

RP REVISIONS.

RC STRAIN-BRISTOL N2;

RA Durbin R.;

RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.

CC -|- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLE).

CC -|- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; Z11115; CAA77460.1; -

DR EMBL; Z22175; CAA77460.1; JOINED.

DR EMBL; Z22175; CAA80131.1; -

DR EMBL; Z11115; CAA80131.1; JOINED.

DR PIR; S15786; S15786.

DR WormPep; ZK637.1; CE06638.

DR InterPro; IPR001066; -

DR Pfam; PF00083; sugar_ttr; 1.

DR PROSITE; PS00216; SUGAR_TRANSPORT_1; FALSE_NEG.

DR PROSITE; PS00217; SUGAR_TRANSPORT_2; FALSE_NEG.

KW Hypothetical protein; Duplication; Transmembrane; Transport.

FT TRANSMEM 86 106 POTENTIAL.

FT TRANSMEM 122 142 POTENTIAL.

FT TRANSMEM 158 178 POTENTIAL.

FT TRANSMEM 238 258 POTENTIAL.

FT TRANSMEM 320 340 POTENTIAL.

FT TRANSMEM 373 393 POTENTIAL.

FT TRANSMEM 411 431 POTENTIAL.

FT TRANSMEM 482 502 POTENTIAL.

SQ SEQUENCE 529 AA; 58317 MW; 8D2FF4CBA15ECD2D CRC64;

Query Match 12.3%; Score 355.5; DB 1; Length 529;

Best Local Similarity 26.5%; Pred. No. 5e-16;

Matches 140; Conservative 97; Mismatches 186; Indels 105; Gaps 24;

QY 52 VPDAAANLSSAWRNHTVPLRLRDGREVPHSCRRYRLATIANFSALGLEP--GRDVD----- 104

Db 10 VLEASNLEAYVLTAKQLKEIRHVGDDP-AVRYSLDDORTELG-EPTDQRSPDSEKTF 67

QY 105 -----LGQLEQF-SCLDGWESQD-----VYLSTI-----VTEWNLVCEDDWKAPLT 145

Db 68 TVDEAVEALGGRFQLKLSILGTMANMADAMEMMLLSLSPALACEWGI---SSVQQALV 124

QY 146 ISLFFVGVLLGSFISGQSDRFG-RKNVLVFT-----MGMTGFS---FLQIFSKNFEMF 196

Db 125 TTCVFSGMMLSSTFWGKICDRFGRKGLTFTSLVACIMGVISGSPHFYVLLFRGLTGF 184

QY 197 VLVFLVLMGQISNYVAAFV--LGTEILGKSVRIETSLGVCIFYAGYVMVLPLEFAVFTV 254

Db 185 -----GIGVFQSVTLYAEFLPTQRAKCVVLIIS-----FWAIGAVFEALLAYFVM 231

QY 255 D---WRMLLVALTMP-GVLCAVLMWFIPESPRWLISQGRFEEAEVIRKAAKANGIVWPS 310

Db 232 ESFGWRALMFLSLPLGIFAVASEW-LPESARFDMASGHPERALETLQAAARNRVQLPT 290

QY 311 TIFDPSELQDLSSKKQOSHNLDL--LRTNIRMTVIMSIMLWMTISVGFGLSLDTPNL 368

Db 291 G-----RLVSTKAGSESRGDIANLLSPDLRKTILLWCIIWATAFSYGVMVLEFTVL 343

QY 369 -----HGDIFFVN-----CFLSAMVEVPAYVLAWLILLOYLPRYSNA 404

Db 344 FQSHDECHGGLFSNGTQMEVCQPLTRSDYFDLLSTTLAEFPGLIITVLIIEWFGKKTMA 403

QY 405 T--ALFLGGVSLFLFMQLVPPDLYLATVLMVGKFGVTAFAFWMVYVYTAELPTVTVRNMG 462

Db 404 LEVAVFAIFTFLLYFCILD----RFTVTVLIFVARAFISGAFQACAYVTVPEVYPTTLRAVG 459

QY 463 VGVSTASRLGSLSPYVYVGLAYDFLPYILMGSLTILTAITLFLP 510

Db 460 LGTCSAMARIGAIVA-----SEKSLPIGIYGTAILGLIASLIP 501

RESULT 7

YGCS_ECOLI STANDARD; PRT; 445 AA.

AC Q46909;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 01-NOV-1997 (Rel. 35, Last annotation update)

DE HYPOTHETICAL METABOLITE TRANSPORT PROTEIN IN CYSJ-ENO INTERGENIC

DE REGION.

GN YGCS.

OS Escherichia coli.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Escherichia.

OX NCBI_TaxID=562;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=K12 / MCL1655;

RX MEDLINE=97426617; PubMed=9278503;

RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,

RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,

RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,

RA Mau B., Shao Y.;

RT "The complete genome sequence of Escherichia coli K-12.";

RL Science 277:1453-1474(1997).

CC -|- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE

CC (POTENTIAL).

CC -|- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```
CC -----
CC EMBL: U29579; AAA69281.1; ALT_INIT.
CC EMBL: AE000360; AAC75813.1; ALT_INIT.
CC EcoGene: EG13126.ygcS.
CC InterPro: IPR001066; -.
CC Pfam: PF00083; sugar_tr; 1.
CC PROSITE: PS00216; SUGAR_TRANSPORT_1; FALSE_NEG.
CC PROSITE: PS00217; SUGAR_TRANSPORT_2; 1.
CC KW Hypothetical protein; Transmembrane; Inner membrane.
CC FT TRANSMEM 23 43 POTENTIAL.
CC FT TRANSMEM 57 77 POTENTIAL.
CC FT TRANSMEM 86 106 POTENTIAL.
CC FT TRANSMEM 115 135 POTENTIAL.
CC FT TRANSMEM 143 163 POTENTIAL.
CC FT TRANSMEM 176 196 POTENTIAL.
CC FT TRANSMEM 254 274 POTENTIAL.
CC FT TRANSMEM 287 307 POTENTIAL.
CC FT TRANSMEM 312 332 POTENTIAL.
CC FT TRANSMEM 338 358 POTENTIAL.
CC FT TRANSMEM 370 390 POTENTIAL.
CC FT TRANSMEM 401 421 POTENTIAL.
CC SQ SEQUENCE 445 AA; 48234 MW; B59E452721B15774 CRC64;

Query Match 10.7%; Score 308; DB 1; Length 445;
Best Local Similarity 26.6%; Pred. No. 4.7e-13;
Matches 102; Conservative 79; Mismatches 168; Indels 34; Gaps 11;

QY 151 VGVLLSFTSGQLSDRFGRKNVLFVTMGMTGFSFLQFSKFNFMFVFLVFLVGMQISN 210
DB 65 LGLFLGSLVGLWISDRHGKIFTSFLITLASFLQFFATTPHELIGRLILIGLGGD 124
QY 211 YVAAFVLGTGILCKSVRIIFSTLGV-CIFYAFGYMVLPLFA-YFIRD----WRMLLVA 264
DB 125 YSVCHVLLAAEFSPRRHGIL--LGAFSVVTWVGVLASIAAGHHFISENPEARWLLASAA 182
QY 265 MPGVCLVALMWFIPESPRWLISGRFEAEVIRKAAKANGIVPSTIPDPSEL--QDLS 322
DB 183 LPALLITLLNWGTPSPRWLLRQGEAAEHAIVHR-----YFGPHVLLGDEVV 230
QY 323 SKKQSHNILLRTNIRNVTIMSLMWTISVGYEGLSDLPNLHGDFVNCFLSAMV 382
DB 231 TATHKHKHTLFSSRYR---RTAFNSVFEVCLVPIWFIYTWLPTAQIGLEDALTA 287
QY 383 EVPAYVLAWLQLYLPRLRYSMATLFLGGSVLFF-----MQLVPPDLXYLATVLMVGKF 437
DB 288 MNLALLIVGALLG-LVLTLLAHRKFLGSLFLLAATLVMACLPSGSSITLLLVLFST 346
QY 438 GVTAAFSMVYVYTAELYPTVVRNMGVSVSTASRLGSLSPYFV--YLGAYDRFLPYILM 495
DB 347 TISAVSNLGVLPASFPPTDIRSLGVGFATMSRLGAASVSTGLLPVLAQWGMQVTLILL 406
QY 496 GSLTILTAILT-LFLPESFGTPL 517
DB 407 ATVLLGVFVVTWLPWAPETKALPL 429

RESULT 8
YTI3_CAEL STANDARD; PRT; 435 AA.
AC Q10917;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 01-OCT-1996 (Rel. 34, Last annotation update)
DE HYPOTHETICAL 48.6 KDA PROTEIN B0252.3 IN CHROMOSOME II.
GN B0252.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Du 2., Waterston R.;
RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
CC -|- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```



```

RESULT 12
SV2_RAT
ID SV2_RAT STANDARD; PRT; 742 AA.
AC Q02563;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE SYNAPTIC VESICLE PROTEIN 2 (SV2).
GN SV2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.; AND SEQUENCE OF 1-40.
RX TISSUE=Brain;
RX MEDLINE=92390722; PubMed=1519064;
RA Bajjalieh S.M., Peterson K., Shingal R., Scheller R.H.;
RT "SV2, a brain synaptic vesicle protein homologous to bacterial
transporters";
RL Science 257:1271-1273(1992).
CC -1- FUNCTION: MAY ACT AS A VESICLE-LOCALIZED NEUROTRANSMITTER
TRANSPORTER OR MAY FUNCTION AS AN ION TRANSPORTER OR CHANNEL.
CC -1- SUBCELLULAR LOCATION: SYNAPTIC VESICLE.
CC -1- TISSUE SPECIFICITY: NEURAL AND ENDOCRINE CELLS OF BRAIN AND
SPINAL CORD.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L05435; AAA42188.1; -.
KW Pfam; PF00083; sugar_tr; 1.
KW Synapse; Nerve; Glycoprotein; Neurotransmitter transport;
Transmembrane.
FT DOMAIN 1 163 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 164 182 POTENTIAL.
FT DOMAIN 183 204 LUMENAL (POTENTIAL).
FT TRANSMEM 205 225 POTENTIAL.
FT DOMAIN 226 242 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 243 260 POTENTIAL.
FT DOMAIN 261 262 LUMENAL (POTENTIAL).
FT TRANSMEM 263 281 POTENTIAL.
FT DOMAIN 282 294 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 295 322 POTENTIAL.
FT DOMAIN 323 334 LUMENAL (POTENTIAL).
FT TRANSMEM 335 355 POTENTIAL.
FT DOMAIN 356 445 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 446 469 POTENTIAL.
FT DOMAIN 470 592 LUMENAL (POTENTIAL).
FT TRANSMEM 593 611 POTENTIAL.
FT DOMAIN 612 626 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 627 647 POTENTIAL.
FT DOMAIN 648 649 LUMENAL (POTENTIAL).
FT TRANSMEM 650 669 POTENTIAL.
FT DOMAIN 670 694 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 695 711 POTENTIAL.
FT DOMAIN 712 712 LUMENAL (POTENTIAL).
FT TRANSMEM 713 731 POTENTIAL.
FT DOMAIN 732 742 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 498 498 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 548 548 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 573 573 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 742 AA; 82705 MW; 565DE7EF2929D5B CRC64;

```

Query Match

Best Local Similarity

9.1%; Score 262.5; DB 1; Length 742;

17.5%; Pred. No. 6.6e-10;

```

Matches 115; Conservative 103; Mismatches 173; Indels 267; Gaps 21;
QY 96 GLPEGRDVLGQLEQSCLDGWEFSODVYLSTIVTE-----WNL----- 134
DB 125 GLSDGEGPPGGRGAORRKRDLAAQ--YETILRCGHRGFWTLTYFVLGLALMADGVE 182
QY 135 -----VCEDDWKAPLTISLFFVGYLLGSFSGQSDRFGKRNKLVFTVMQMG 180
DB 183 VFVVGFLPSAEKDMCLSDSNKMGMLGLIVLGMVGAFLMGGLADRLGRROCLLSLSV 242
QY 181 TGSFSLQIFSKNFEMFVLVFLVGMQGISNYVAAFLVGLTEILGKSVRIIFSLG----- 234
DB 243 SVFAFFSSFOGGTGTFLFCLLSGVG-----IGGSIPIVFTSFSEFLAQE 287
QY 235 -----VCIFYAFG-----YMLPLF-----AYFIQWRMLLVALTMPGVLC 270
DB 288 KRGEHLWLCMFWMVIGVYAAAMAWAIIPIHYGWSFOMGSAYQFHSWRVFLVFAFSPVFA 347
QY 271 VALWVFIPESPRWLISQGRFEAEVIRKA---AKANGIVVPSTIFDPSSELDLSSKKQ 326
DB 348 IGALTTOPESPREFLENGKHDEAMVILKQVHDTNMRAG--HPERVSVTHIKTI----H 401
QY 327 QSHNILD-----LRTWNI-----RMVTIMSIMLWMTISVGY 358
DB 402 QEDLELIEIQSDTGTWQRCVGRALSLGGQWGNFLSCFSPRYRITLMMGMVWFTMSFSY 461
QY 359 FGLSLDTPNL-----FLSANVEVAYVLAWLILLOYLPRRYSMATALELGSVL-- 414
DB 462 YGLTVWPPDMIRHLQAVDVAARTKVPGERVHVTFTLENOIHRGGQYFNDKFIQLRL 521
QY 369 -----HGDIYFNC----- 376
DB 522 KVSFEDSLFECEYFEDVTSSNTFFRNCFTINTVFYNTDLFEYKFNLSRLVNSTFLHNKE 581
QY 377 -----FLSANVEVAYVLAWLILLOYLPRRYSMATALELGSVL-- 414
DB 582 GCPDVTGTGEGAYMYVFSFGLTAVLPGNIVSALLMDKIGRLMLA-----GSSVLSC 636
QY 415 ---LFMQLVPPDLYLATVLVVMVGKFGVT-AAFSMYVYVYTAELYPTVVRNMGVGSSTAS 470
DB 637 VSCFFLSFGNSSESAMIA-LLCLEFG--GVSIASWALDVLTVELYPSDKRTTAFGLNALC 693
QY 471 RLGSIL--SPYFVLGAYDRFLPYILMGSLTILTLTLFLPESFGTPLPOTIDQMLR 526
DB 694 KLAALVGISITSFVG-----ITKAAPILFASALGSLALKLPETRGVOLQ 742
RESULT 13
YFIG_BACSU
ID YFIG_BACSU STANDARD; PRT; 482 AA.
AC P34723;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE HYPOTHETICAL METABOLITE TRANSPORT PROTEIN IN GLVBC 3'REGION.
GN YFIG.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA MEDLINE=96262713; PubMed=8704981;
RA Yamamoto H., Uchiyama S., Fajar A.N., Ogasawara N., Sekiguchi J.;
RT "Determination of a 12 kb nucleotide sequence around the 76 degrees
RT region of the Bacillus subtilis chromosome.";
RL Microbiology 142:1417-1421(1996).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLE).
CC -1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -

```


CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

 DR EMBL; D50543; BAA09111.1; -;
 DR EMBL; Z99108; CAB12655.1; -;
 DR Subtilist; BG11854; yfig.
 DR InterPro; IPR001066; -;
 DR Pfam; PF00083; sugar_tr; 1.
 DR PRINTS; PS00171; SUGRTNSPORT
 DR PROSITE; PS00216; SUGAR_TRANSPORT_1; FALSE_NEG.
 DR PROSITE; PS00217; SUGAR_TRANSPORT_2; 1.
 KW Hypothetical protein; Transp; Transmembrane.
 FT TRANSMEM 30 50 POTENTIAL.
 FT TRANSMEM 60 80 POTENTIAL.
 FT TRANSMEM 93 113 POTENTIAL.
 FT TRANSMEM 121 141 POTENTIAL.
 FT TRANSMEM 156 176 POTENTIAL.
 FT TRANSMEM 185 205 POTENTIAL.
 FT TRANSMEM 264 284 POTENTIAL.
 FT TRANSMEM 302 322 POTENTIAL.
 FT TRANSMEM 332 352 POTENTIAL.
 FT TRANSMEM 353 373 POTENTIAL.
 FT TRANSMEM 401 421 POTENTIAL.
 FT TRANSMEM 424 444 POTENTIAL.
 SQ SEQUENCE 482 AA; 52756 MW; 90851C4F4C48EE01 CRC64;

Query Match 8.9%; Score: 256; DB 1; Length 482;

Best Local Similarity 23.9%; Pred. No. 1.le-09;
 Matches 105; Conservative 83; Mismatches 167; Indels 84; Gaps 19;

QY 144 LRTSLFFVGLSGFTSGOLSDRGKKN-----VLFV--TMGMOTGFSFLQIFSKNPFM 195
 DB 64 LVASSLLGAAFGMEGGRLSDRHGRRTILYLALLFLAATLGC-----TFSPNASV 115
 QY 196 FVFLVFLVGM--GQISNYVAAVFLGTEILGKSVRI-----FSTLGVCIFVAFGYMVLPL 248
 DB 116 MIAFPELLGLAVGCASVTPTFLAELSPAERRGRIVTQNELMIVIGQLLAYTFNAIGST 175
 QY 249 FAYFIDRWMLVALTMCPLVCVALWW---FTIPSPRLISOGFEAEVLIIRKAAKANG 305
 DB 176 MGESANWRYMLVIATLPVAV---LWFGMLVPSRPLAAKGRMGDALVLRQIRE--- 229
 QY 306 IVPSTIFDPSELQDL-----SSKKQOSHNIIDLLRTWNIRMTI---MSIMLWMT 353
 DB 230 -----DSQAQOEKEIKHAIEGTAKKAGFH---DFQEPWIRILFIGIGIAIVQGIT 278
 QY 354 --ISVGYFGLSL-----DTNPLHGDIFVNCFLSAMVEVPAYVLAWLLQLYLPFR--- 400
 DB 279 GVNSIMYGYTEILREAGFQTEAALIGNI-----ANGVISVIAVIFGIWLLGKVRPRMLI 333
 QY 401 ---YSMATALFGLGSLVLLFPMQLPDPDLYLATVLMVGRFGVTAAFSMVYVYTAELYPV 457
 DB 334 IGOIGTWATLLIGILUSIVLEGTALPYVLSLTILFLAQFOATSTVWMLSEIFPMH 393
 QY 458 VRNMGVGVSS---TASRLGSILSPYFV-YLGAYDRFLPYILMGSITILITAILFLPES 512
 DB 394 VRGLGMSITFCFLWTANFLIGFTFPILLNHIGMSATFFIFVAMNLAAIL--FVKYVPET 451
 QY 513 FGTPPLDPTDQMLRVKGMK 531
 DB 452 KGRSL-EQLEHSPFOYGR 469

RESULT 14

YDJE_ECOLI

ID YDJE_ECOLI STANDARD; PRT; 452 AA.

AC P38055; p77244;

DT 01-OCT-1994 (Rel. 30, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE HYPOTHETICAL METABOLITE TRANSPORT PROTEIN IN ANSA-CAPA INTERGENIC
 GN REGION.
 GN YDJE.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OC NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX Aiba H., Baba T., Fujita K., Hayashi K., Honjo A., Horiuchi T.,
 RA Ikemoto K., Inada T., Isono K., Isono S., Itoh T., Kanai K., Kasai H.,
 RA Kashimoto K., Kim S., Kimura S., Kitagawa M., Kitakawa M., Makino K.,
 RA Masuda S., Miki T., Mizobuchi K., Mori H., Motomura K., Nakamura Y.,
 RA Nashimoto H., Nishio Y., Oshima T., Saito N., Sampaio G., Seki Y.,
 RA Tagami H., Takemoto K., Wada C., Yamamoto Y., Yano M.;
 RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 360-452 FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=89357501; PubMed=2670682;
 RA Jellstrom P.G., Bezjak D.A., Jennings M.P., Beacham I.R.;
 RT "Structure and expression in Escherichia coli K-12 of the
 RL L-asparaginase I-encoding *ansA* gene and its flanking regions.";
 RN Gene 78:37-46(1989).
 RP IDENTIFICATION.
 RX MEDLINE=95075659; PubMed=7984428;
 RA Borodovsky M., Rudd K.E., Koonin E.V.;
 RT "Intrinsic and extrinsic approaches for detecting genes in a
 RL bacterial genome.";
 CC Nucleic Acids Res. 22:4756-4767(1994).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
 CC (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

 CC EMBL; AE000272; AAC74839.1; -;
 CC EMBL; D90820; BAA15560.1; -;
 CC EMBL; D90821; BAA15567.1; -;
 CC EMBL; M26934; -; NOT_ANNOTATED_CDS.
 CC EcoGene; EG12369; ydJE.
 CC InterPro; IPR001066; -;
 CC Pfam; PF00083; sugar_tr; 1.
 DR PROSITE; PS00216; SUGAR_TRANSPORT_1; FALSE_NEG.
 DR PROSITE; PS00217; SUGAR_TRANSPORT_2; FALSE_NEG.
 KW Hypothetical protein; Transp; Transmembrane; Inner membrane.
 FT TRANSMEM 23 43 POTENTIAL.
 FT TRANSMEM 60 80 POTENTIAL.
 FT TRANSMEM 96 116 POTENTIAL.
 FT TRANSMEM 119 139 POTENTIAL.
 FT TRANSMEM 167 187 POTENTIAL.
 FT TRANSMEM 264 284 POTENTIAL.
 FT TRANSMEM 299 319 POTENTIAL.
 FT TRANSMEM 327 347 POTENTIAL.

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.
OM protein - protein search, using sw model
Run on: August 16, 2001, 14:03:35 ; Search time 44.83 Seconds
(without alignments)
1643.852 Million cell updates/sec

Title: US-09-521-195-3
Perfect score: 2883
Sequence: 1 MRDYDEVTAFLGEGWGFQRL.....HTRMLKDGQRPILKSTAF 557

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTRMBL_16.*
1: sp.archaea.*
2: sp.bacteria.*
3: sp.fungi.*
4: sp.human.*
5: sp.invertebrate.*
6: sp.mammal.*
7: sp.mhc.*
8: sp.organelle.*
9: sp.phage.*
10: sp.plant.*
11: sp.rodent.*
12: sp.unclassified.*
13: sp.vertebrate.*
14: sp.virus.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2295.5	79.6	564	11 Q9WTN6	Q9wt6 mus musculus
2	2227	77.2	551	4 Q9R015	Q9R015 homo sapien
3	2218	76.9	551	4 O14546	O14546 homo sapien
4	2157	74.8	553	11 Q9R141	Q9R141 rattus norv
5	2148	74.5	553	11 Q9Z306	Q9Z306 mus musculus
6	885	30.7	548	5 Q9VCA2	Q9vca2 drosophila
7	881	30.6	548	5 O01384	O01384 drosophila
8	816	28.3	567	5 Q9VCA3	Q9vca3 drosophila
9	777	27.0	568	5 Q9U539	Q9U539 caenorhabdi
10	777	27.0	576	5 O02270	O02270 caenorhabdi
11	753.5	26.1	561	5 Q9V6L6	Q9v616 drosophila
12	740	25.7	554	4 Q9N0D4	Q9nqd4 homo sapien
13	738.5	25.6	556	4 O75751	O75751 homo sapien
14	738	25.6	554	4 O15395	O15395 homo sapien
15	734	25.5	554	4 O15245	O15245 homo sapien
16	733	25.4	593	11 P70485	P70485 rattus norv
17	728	25.3	593	11 P97558	P97558 rattus norv
18	727	25.2	555	4 O15244	O15244 homo sapien
19	724	25.1	555	11 Q9R0W2	Q9R0W2 rattus norv

```
077504 oryctolagus
063089 rattus norv
09vik2 drosophila
070577 mus musculu
002713 sus scrofa
063314 rattus norv
008966 mus musculu
09rlq4 mus musculu
09h2w5 homo sapien
09y226 homo sapien
09wtw5 mus musculu
09y694 homo sapien
088446 rattus norv
057379 pseudopleur
014567 homo sapien
088909 mus musculu
09ri17 rattus norv
09vex8 drosophila
09v6h5 drosophila
035956 rattus norv
09vnx8 drosophila
061185 mus musculu
095742 homo sapien
09vey2 drosophila
054778 mus musculu
09tsy7 oryctolagus
```

```
20 723.5 25.1 554 6 077504
21 718 24.9 556 11 063089
22 716.5 24.9 674 5 09vik2
23 714.5 24.8 553 11 070577
24 711 24.7 554 6 002713
25 710 24.6 535 11 063314
26 709 24.6 556 11 008966
27 704 24.4 556 11 09rlq4
28 703.5 24.4 539 4 09h2w5
29 703 24.4 551 4 09y226
30 690.5 24.0 551 11 09wtw5
31 690 23.9 548 4 09y694
32 681.5 23.6 551 11 088446
33 675.5 23.4 562 13 057379
34 660.5 22.9 456 4 014567
35 660.5 22.9 537 11 088909
36 656.5 22.8 536 11 09ri17
37 651 22.6 557 5 09vex8
38 650 22.5 538 5 09v6h5
39 643.5 22.3 551 11 035956
40 638.5 22.1 762 5 09vnx8
41 638 22.1 545 11 061185
42 628 21.8 563 4 095742
43 621 21.5 569 5 09vey2
44 620 21.5 553 11 054778
45 611.5 21.2 551 6 09tsy7
```

ALIGNMENTS

RESULT 1

```
Q9WTN6 Q9WTN6 PRELIMINARY; PRT; 564 AA.
ID Q9WTN6;
AC Q9WTN6;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE OCTN3.
GN OCTN3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=TESTIS;
RA Nezu J.;
RT "Mouse OCTN3 - a novel OCTN transporter family protein.";
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB018436; BAA78343.1; -.
DR InterPro; IPR001066; -.
DR Pfam; PF00083; sugar_tr; 1.
DR PROSITE; PS00216; SUGAR_TRANSPORT_1; UNKNOWN_1.
SQ SEQUENCE 564 AA; 63320 MW; C37FDC6395DAD01 CRC64;
```

Query Match 79.6%; Score 2295.5; DB 11; Length 564;
Best Local Similarity 80.6%; Pred. No. 1.4e-137;
Matches 435; Conservative 54; Mismatches 48; Indels 3; Gaps 1;

```
QY 1 MRDYDEVTAFLGEGWGFQRLIFFLLSASIIPIGFTGLSSVELIATPEHRCVDPDAANLSS 60
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 MRDYDEVTAFLGEGWGFQRLIFFLLSASIIPIGFTGLSSVELIATPEHRCVDPDAANLSS 60
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 61 AWRNHTVPLRLRDGVEPHSCRRYRLATIANFSAFLGPRDVLGLQEQSLDGEWFFS 120
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 AWRNHSIPMETKDGFEVQKCRRYRLATIANFSELGLPGRDVLGLQEQENCLDGEWD 120
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 121 QDVYLSTIVTEWNLVCEDDWKAPLITISLFFVGLVGLSGISQSLDRFGKKNVLFVTMGWQ 180
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 KDIFLSTIVTEWDLCKDWDKAPLITTSFFYVGLVGLSGISQSLDRFGKKNLFLTMAMH 180
```

```
QY 181 TGSFLQIFSKNFEMFVFLVLMGQISNYVAAFVLGTIELGKSVRIIFSTLGVCIYA 240
Db 181 TGSFIQVFSVNFEMFTLLTYLVGMGHISNYVAAFVLGTIELGKSVRIIFAILGVCIFA 240
QY 241 FGYMVLPLFAFYFIRDMRLMLVALTMFGVLCVALWPFIPESPRWLISQGRPEAEVIIRKA 300
Db 241 FGYMVLPLFAFYFIRDMRLMLVALTMFGVLCVALWPFIPESPRWLISQGRPEAEVIIRKA 300
QY 301 AKANGIVVPSTIFDPSE--LQDLSSKKQSHNLDLLRTWNRMTYIMSLMWTISVG 357
Db 301 AKANGIVVPSTIFDPSE--LQDLSSKKQSHNLDLLRTWNRMTYIMSLMWTISVG 357
QY 358 YFGLSLDTPNLHGDIYVNCFLSAMVEVPAYVLAWLLQYLPRRYSMATLFLGGSVLLFM 417
Db 358 YFGLSLDTPNLHGDIYVNCFLSAMVEVPAYVLAWLLQYLPRRYSMATLFLGGSVLLFM 417
QY 418 QLYPPDLIYLATVLMVGKFGVTAFAFVYVYTAELPTVVRNMGVGSSTASRLGSIIS 477
Db 418 QLYPPDLIYLATVLMVGKFGVTAFAFVYVYTAELPTVVRNMGVGSSTASRLGSIIS 477
QY 478 PYFVILGAYDRFLPYILMGSLTILTAITLFLPESFGTLPDTIDQMLRVKGMKHKTKPS 537
Db 478 PYFVILGAYDRFLPYILMGSLTILTAITLFLPESFGTLPDTIDQMLRVKGMKHKTKPS 537
QY 481 PYFVILGAYDRFLPYILMGSLTILTAITLFLPESFGTLPDTIDQMLRVKGMKHKTKPS 540
Db 481 PYFVILGAYDRFLPYILMGSLTILTAITLFLPESFGTLPDTIDQMLRVKGMKHKTKPS 540
RESULT 2
Q9H015 PRELIMINARY; PRT; 551 AA.
ID Q9H015
AC Q9H015
DT 01-JAN-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE INTEGRAL MEMBRANE TRANSPORT PROTEIN.
GN UT2H.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RA Spritzberger F., Gruendemann D., Schoemig E.;
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: Y09881; CAAT1007.1;
SQ SEQUENCE 551 AA; 62155 MW; C827A99AA78C9443 CRC64;
```

Query Match 77.2%; Score 2227; DB 4; Length 551;
Best Local Similarity 76.3%; Pred. No. 3e-133;
Matches 425; Conservative 57; Mismatches 69; Indels 6; Gaps 3;

```
QY 1 MRDYDEVTAFLGEGPQRLIFLLSASIPNGFTGLSSVFLIATPEHRCRVPDANLSS 60
Db 1 MRDYDEVTAFLGEGPQRLIFLLSASIPNGFTGLSSVFLIATPEHRCRVPDANLSS 60
QY 61 AWRNHTVPLRLDGRVPHSCRYRLATIANFSALEGPGRDVLGQLEQESCLDGWEFS 120
Db 61 AWRNHTVPLRLDGRVPHSCRYRLATIANFSALEGPGRDVLGQLEQESCLDGWEFS 120
QY 121 QDYLSTVTEWNNVCEDDWKAPLTISLFFVGVLLGSGTSIGLSRFRGKNVFLVTGMQ 180
Db 121 QDYLSTVTEWNNVCEDDWKAPLTISLFFVGVLLGSGTSIGLSRFRGKNVFLVTGMQ 180
QY 181 TGSFLQIFSKNFEMFVFLVLMGQISNYVAAFVLGTIELGKSVRIIFSTLGVCIYA 240
Db 181 TGSFLQIFSKNFEMFVFLVLMGQISNYVAAFVLGTIELGKSVRIIFSTLGVCIYA 240
QY 241 FGYMVLPLFAFYFIRDMRLMLVALTMFGVLCVALWPFIPESPRWLISQGRPEAEVIIRKA 300
Db 241 FGYMVLPLFAFYFIRDMRLMLVALTMFGVLCVALWPFIPESPRWLISQGRPEAEVIIRKA 300
QY 301 AKANGIVVPSTIFDPSE--LQDLSSKKQSHNLDLLRTWNRMTYIMSLMWTISVG 357
Db 301 AKANGIVVPSTIFDPSE--LQDLSSKKQSHNLDLLRTWNRMTYIMSLMWTISVG 357
QY 358 YFGLSLDTPNLHGDIYVNCFLSAMVEVPAYVLAWLLQYLPRRYSMATLFLGGSVLLFM 417
Db 358 YFGLSLDTPNLHGDIYVNCFLSAMVEVPAYVLAWLLQYLPRRYSMATLFLGGSVLLFM 417
QY 418 QLYPPDLIYLATVLMVGKFGVTAFAFVYVYTAELPTVVRNMGVGSSTASRLGSIIS 477
Db 418 QLYPPDLIYLATVLMVGKFGVTAFAFVYVYTAELPTVVRNMGVGSSTASRLGSIIS 477
QY 478 PYFVILGAYDRFLPYILMGSLTILTAITLFLPESFGTLPDTIDQMLRVKGMKHKTKPS 537
Db 478 PYFVILGAYDRFLPYILMGSLTILTAITLFLPESFGTLPDTIDQMLRVKGMKHKTKPS 537
QY 481 PYFVILGAYDRFLPYILMGSLTILTAITLFLPESFGTLPDTIDQMLRVKGMKHKTKPS 540
Db 481 PYFVILGAYDRFLPYILMGSLTILTAITLFLPESFGTLPDTIDQMLRVKGMKHKTKPS 540
```

```
QY 361 LSLDTPNLHGDIYVNCFLSAMVEVPAYVLAWLLQYLPRRYSMATLFLGGSVLLFMQ 420
Db 361 LSLDTPNLHGDIYVNCFLSAMVEVPAYVLAWLLQYLPRRYSMATLFLGGSVLLFMQ 420
QY 421 PDLYLATVLMVGKFGVTAFAFVYVYTAELPTVVRNMGVGSSTASRLGSIIS 480
Db 421 PDLYLATVLMVGKFGVTAFAFVYVYTAELPTVVRNMGVGSSTASRLGSIIS 480
QY 481 VYLGAIDRFLPYILMGSLTILTAITLFLPESFGTLPDTIDQMLRVKGMKHKTKPS 540
Db 481 VYLGAIDRFLPYILMGSLTILTAITLFLPESFGTLPDTIDQMLRVKGMKHKTKPS 540
QY 541 MKDQOERPTILKSTAF 557
Db 541 MKDQOERPTILKSTAF 557
QY 551 MKDQOERPTILKSTAF 557
Db 551 MKDQOERPTILKSTAF 557
RESULT 3
O14546 PRELIMINARY; PRT; 551 AA.
ID O14546
AC O14546
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE POLYSPECIFIC ORAGANIC CATION TRANSPORTER.
GN OCTN1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RA Tamai I., Yabuuchi H., Nezu J., Sai Y., Oku A., Shimane M., Tsuji A.;
RL "Cloning and characterization of a novel human pH-dependent organic
RL cation transporter, OCTN1."
RL FEBS Lett. 419:107-111(1997).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.
DR EMBL: AB007448; BAA23356.1; -.
DR InterPro; IPR001066; -.
DR Pfam; PF00083; sugar_tr; 1.
DR PROSITE; PS00216; SUGAR_TRANSPORT_1; UNKNOWN_1.
KW Transmembrane.
SQ SEQUENCE 551 AA; 62176 MW; F5903421C789F60A CRC64;
```

Query Match 76.9%; Score 2218; DB 4; Length 551;
Best Local Similarity 75.9%; Pred. No. 1.1e-132;
Matches 423; Conservative 57; Mismatches 71; Indels 6; Gaps 3;

```
QY 1 MRDYDEVTAFLGEGPQRLIFLLSASIPNGFTGLSSVFLIATPEHRCRVPDANLSS 60
Db 1 MRDYDEVTAFLGEGPQRLIFLLSASIPNGFTGLSSVFLIATPEHRCRVPDANLSS 60
QY 61 AWRNHTVPLRLDGRVPHSCRYRLATIANFSALEGPGRDVLGQLEQESCLDGWEFS 120
Db 61 AWRNHTVPLRLDGRVPHSCRYRLATIANFSALEGPGRDVLGQLEQESCLDGWEFS 120
QY 121 QDYLSTVTEWNNVCEDDWKAPLTISLFFVGVLLGSGTSIGLSRFRGKNVFLVTGMQ 180
Db 121 QDYLSTVTEWNNVCEDDWKAPLTISLFFVGVLLGSGTSIGLSRFRGKNVFLVTGMQ 180
QY 181 TGSFLQIFSKNFEMFVFLVLMGQISNYVAAFVLGTIELGKSVRIIFSTLGVCIYA 240
Db 181 TGSFLQIFSKNFEMFVFLVLMGQISNYVAAFVLGTIELGKSVRIIFSTLGVCIYA 240
QY 241 FGYMVLPLFAFYFIRDMRLMLVALTMFGVLCVALWPFIPESPRWLISQGRPEAEVIIRKA 300
Db 241 FGYMVLPLFAFYFIRDMRLMLVALTMFGVLCVALWPFIPESPRWLISQGRPEAEVIIRKA 300
```

```
QY 301 AKANGIWPSTIFDPSELQDLSSKKQSHNILDLLRTNIRMTIMSLMWTISVGYFG 360
Db 301 AKNNTPNHLGDI FVNCFLSAMVEVPAYVLAWLLQLYLP RRYSMATALFLGSGVLLPMOLV 420
QY 361 LSLDTPNHLGDI FVNCFLSAMVEVPAYVLAWLLQLYLP RRYSMATALFLGSGVLLPMOLV 420
Db 359 LSLDAPNHLGDI FVNCFLSAMVEVPAYVLAWLLQLYLP RRYSMATALFLGSGVLLPMOLV 418
QY 421 PPDLYLATVLMVVGKFGVTAAFSMVYVTAELYP TVVRNMGVSGVSTASRLGSLSPYF 480
Db 419 PVDYFSLGLWLGKFGVTAAFSMVYVTAELYP TVVRNMGVSGVSTASRLGSLSPYF 478
QY 481 VYLGAIDRFLPYILMGSLTILTAITLFLPESFGTLPDPTIDQMLRVKGMKHKRTPSHTR 540
Db 479 VYLGAIDRFLPYILMGSLTILTAITLFLPESFGTLPDPTIDQMLRVKGMKHKRTPSHTR 538
QY 541 MLKDGQERTILKSTAF 557
Db 536 DSMETEENPKVL-ITAF 551

RESULT 4
Q9RI41 PRELIMINARY; PRT; 553 AA.
AC Q9RI41;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DE ORGANIC CATION TRANSPORTER OCTN1.
GN OCTN1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RA Wu X., Wang H., Leibach F.H., Ganapathy V.;
RT "Functional characteristics and tissue distribution pattern of OCTN1,
an organic cation transporter";
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF169831; AAD46922.1;
DR InterPro; IPR001066;
DR Pfam; PF00083; sugar tr; 1.
DR PROSITE; PS00216; SUGAR_TRANSPORT_1; UNKNOWN_1.
SQ SEQUENCE 553 AA; 62362 MW; E26C8155768A14AD CRC64;
```

```
Query Match 74.8%; Score 2157; DB 11; Length 553;
Best Local Similarity 72.7%; Pred. No. 8.1e-129;
Matches 405; Conservative 65; Mismatches 83; Indels 4; Gaps 2;

QY 1 MRDYDEVTAFLGEMGPFQRLIFLLSASIPNGFTGLSSVFLIATPEHRCRVPDAANLSS 60
Db 1 MRDYDEVTAFLGEMGPFQRLIFLLSASIPNGFTGLSSVFLIATPEHRCRVPDAANLSS 60
QY 61 AWRNHTVPLRLRDGRVPHSCRRYRLATIANFSALEPCRDVDLQLEQESCLDGWEYS 120
Db 61 AWRNHTVPLRLRDGRVPHSCRRYRLATIANFSALEPCRDVDLQLEQESCLDGWEYS 120
QY 121 QDVLSTIVTEWNLVCEDDWKAPLTISLFFVGVLLGSGQLSDRFGKKNVLFVTMGQ 180
Db 121 QDVLSTIVTEWNLVCEDDWKAPLTISLFFVGVLLGSGQLSDRFGKKNVLFVTMGQ 180
QY 181 TGFSEFQIFSKNFEMFVFLVGMGQISNYVAAFVLGTEILGKSVRIITFSLGVCFIYA 240
Db 181 TGFSEFQIFSKNFEMFVFLVGMGQISNYVAAFVLGTEILGKSVRIITFSLGVCFIYA 240
QY 241 FGYWVLPFAFIRDMRMMLVATMPGVLCVALLWVFPESPRWLSIQGRFEAEVLIIRKA 300
Db 241 FGYWVLPFAFIRDMRMMLVATMPGVLCVALLWVFPESPRWLSIQGRFEAEVLIIRKA 300
QY 301 AKANGIWPSTIFDPSELQDLSSKKQSHNILDLLRTNIRMTIMSLMWTISVGYFG 360
Db 301 AKANGIWPSTIFDPSELQDLSSKKQSHNILDLLRTNIRMTIMSLMWTISVGYFG 360
```

```
Db 301 AKNGIMAPAVIEDPLELQELNSLKQKVFILDLFKTRNIATITVMSVLMWLTISVGYFA 360
QY 361 LSLDTPNHLGDI FVNCFLSAMVEVPAYVLAWLLQLYLP RRYSMATALFLGSGVLLPMOLV 420
Db 361 LSLNPNHLGDI FVNCFLSAMVEVPAYVLAWLLQLYLP RRYSMATALFLGSGVLLPMOLV 420
QY 421 PPDLYLATVLMVVGKFGVTAAFSMVYVTAELYP TVVRNMGVSGVSTASRLGSLSPYF 480
Db 421 PEDYFSLGLWLGKFGVTAAFSMVYVTAELYP TVVRNMGVSGVSTASRLGSLSPYF 480
QY 481 VYLGAIDRFLPYILMGSLTILTAITLFLPESFGTLPDPTIDQMLRVKGMKHKRTPSHTR 540
Db 481 VYLGAIDRFLPYILMGSLTILTAITLFLPESFGTLPDPTIDQMLRVKGMKHKRTPSHTR 537
QY 541 MLKDGQERTILKSTAF 557
Db 538 VSMDEENPKVL-ITAF 553

RESULT 5
Q9Z306 PRELIMINARY; PRT; 553 AA.
AC Q9Z306;
DT 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
DE ORGANIC CATION TRANSPORTER.
GN OCTN1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RA Nezu J.;
RT "Mouse OCTN1: Polyspecific organic cation transporter";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.
DR EMBL; AB016257; BAA36626.1;
DR InterPro; IPR001066;
DR Pfam; PF00083; sugar tr; 1.
DR PROSITE; PS00216; SUGAR_TRANSPORT_1; UNKNOWN_1.
KW Transmembrane.
SQ SEQUENCE 553 AA; 62290 MW; C4D66BC061398653 CRC64;
```

```
Query Match 74.5%; Score 2148; DB 11; Length 553;
Best Local Similarity 72.2%; Pred. No. 3e-128;
Matches 402; Conservative 67; Mismatches 84; Indels 4; Gaps 2;

QY 1 MRDYDEVTAFLGEMGPFQRLIFLLSASIPNGFTGLSSVFLIATPEHRCRVPDAANLSS 60
Db 1 MRDYDEVTAFLGEMGPFQRLIFLLSASIPNGFTGLSSVFLIATPEHRCRVPDAANLSS 60
QY 61 AWRNHTVPLRLRDGRVPHSCRRYRLATIANFSALEPCRDVDLQLEQESCLDGWEYS 120
Db 61 AWRNHTVPLRLRDGRVPHSCRRYRLATIANFSALEPCRDVDLQLEQESCLDGWEYS 120
QY 121 QDVLSTIVTEWNLVCEDDWKAPLTISLFFVGVLLGSGQLSDRFGKKNVLFVTMGQ 180
Db 121 QDVLSTIVTEWNLVCEDDWKAPLTISLFFVGVLLGSGQLSDRFGKKNVLFVTMGQ 180
QY 181 TGFSEFQIFSKNFEMFVFLVGMGQISNYVAAFVLGTEILGKSVRIITFSLGVCFIYA 240
Db 181 TGFSEFQIFSKNFEMFVFLVGMGQISNYVAAFVLGTEILGKSVRIITFSLGVCFIYA 240
QY 241 FGYWVLPFAFIRDMRMMLVATMPGVLCVALLWVFPESPRWLSIQGRFEAEVLIIRKA 300
Db 241 FGYWVLPFAFIRDMRMMLVATMPGVLCVALLWVFPESPRWLSIQGRFEAEVLIIRKA 300
QY 301 AKANGIWPSTIFDPSELQDLSSKKQSHNILDLLRTNIRMTIMSLMWTISVGYFG 360
Db 301 AKANGIWPSTIFDPSELQDLSSKKQSHNILDLLRTNIRMTIMSLMWTISVGYFG 360
```

```

Db 301 AKNSIVAPAGIEPDLQELNSLKQKVILDLFRTRNIATTVMVWMLWMLTSVGIFA 360
QY 361 LSLDTNHLGDIYVNCFLSAMVEVPAYVLAWLILLOVLPYRYSNATFALFGGSSVLFWMQVLY 420
Db 361 LSNVNPVNLGDIYVNCFLSGLIEVPAYFTAWLLRLTPRYIAGVLFWGGVLLIQVY 420
QY 421 PPDLXYLATVLMVVGKFGVTAAFSMVYVTAELPTVVRNMGMVGSSTASRLGSLSPYF 480
Db 421 PEDYNEVSGIWMVGRFGITSAFSMLYVTAELPTVVRNMVAVGITSASRVSIIAPYF 480
QY 481 VYLGAVDRFLPYLMGSLTILTAFLTLFLPESFGTLPDPTIDOMLRVKGMKHKRTPSHTR 540
Db 481 VYLGAVNRLPYLMGSLVLIIGITLFFPESFGVTLPENLEQMKVGRFCGK---KST 537
QY 541 MKDGOERPTILKSTAF 557
Db 538 VSDREESPKVL-ITAF 553

RESULT 6
Q9VCA2 PRELIMINARY; PRT; 548 AA.
AC Q9VCA2;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE ORCT PROTEIN.
GN ORCT OR CG6331.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
NCBI_TaxID=7227;
[1]
SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers J.-H.C., Blazei R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abill J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Jallali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,

```

```

RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
CC -|- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
DR EMBL; AE003747; AAF56271.1; -.
CC -|- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.
DR FlyBase; FBgn0019952; Orct.
DR InterPro; IPR001066; -.
DR Pfam; PF00083; sugar_tr; 1.
KW Transmembrane.
SQ SEQUENCE 548 AA; 61002 MW; 08D7F97599B477AF CRC64;

Query Match 30.7%; Score 885; DB 5; Length 548;
Best Local Similarity 35.1%; Pred. No. 2.le-48;
Matches 198; Conservative 106; Mismatches 206; Indels 54; Gaps 9;

QY 4 YDEVTAFLGEMGPFQRLIFFLLSASIIIPNGFTGLSSVFLIATPEHRCRVP----DAANLS 59
Db 3 YDDVITHLGEFGPYQKRIYYLLCLPAIVCAFHKLAGVFLAKPDRFCALPYENGSIYELS 62
QY 60 SAWRNHTVPLRLDRGVRPHSCRRYRLATTANFSALEGRVDL-----GQL----- 108
Db 63 PHLNLSYPENER-----CSYY-----DVDYTEYLINGSIPRSSN 97
QY 109 QEESCLDGEFSQDVYLSTIVTEWNLVCEDDWKAPLTISLFFVGLGFSIGOLSDFRG 168
Db 98 ETKTC-SSYVYDRSKYLSAVTEWNLVCSRLSLSATSDSLFLMGLVGLSLIFGMSDKLG 156
QY 169 RKNVLFVTMGMTQSFSLQIFSKNFEMFVFLVFLVGMGQISNYVAAPVLTGTEILGKSVRI 228
Db 157 RKPTFFASLVLLQILFGVLAAVAPEYFSYITSRMVGATTSGVFLVAVVALEVMVSSYR- 215
QY 229 IFSTLGVCIYAFGYMVLPLFAYFIRDRMLLVALTWPGVLCVALWVFIPESPRLISQ 288
Db 216 LFAGVAMQMFSGVFMLTAGFAFIHWRWLQITAILPGLLFLCYWIIPEASRWLLMK 275
QY 289 RFEEAEVIRKAANGIVVPSITFD-----PSEQLDLSKKQKQSHNTLDLLRTWNIR 341
Db 276 RKDEAFVIEKAENKENVPEIYELVDVEAEKKQDEMAASQPAATVFDLLRYPNLR 335
QY 342 MVTMTSMLMWTISGVYFGLSLDTPNLHGDFVNCFLSAMVEVPAYVLAWLLOYLPRRY 401
Db 336 RKTLLIFDFNFVNGVYGLSWNTNLGGNQLNFMISGAVEIPGYTLLFTLNRWGRS 395
QY 402 SMATALFLGCVLLFMOLVPPDLYLATVLMVKCKGVTAAFSWVYVYTAELYPTVVRNM 461
Db 396 ILCTMVMAGLSLTLATTFVPSDMNWLIVACAMIGKLAITSYGTIYFSAEQPTVVRNV 455
QY 462 GVGVSSTASRLGSLSPYFVYLGAIDRFLPYILMGSLTILTAFLTLFLPESFGTLPDTI 521
Db 456 GLGASSMVARVGGILAPYLLKLGELINRPLLIICGALSITAGLSLLPPLTNKPNPETI 515
QY 522 DQMLRVKGMKHKRTPSHTRMLKOG 545
Db 516 ED-----GENFGKKPAPQETAEEG 534

RESULT 7
C01384 PRELIMINARY; PRT; 548 AA.
AC C01384;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE PUTATIVE ORGANIC CATION TRANSPORTER.
GN ORCT OR CG6331.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
NCBI_TaxID=7227;
[1]

```


QY 62 WRNHTVPLRLDGRVPHSCRYRLATIANPSALGLEPGRDVLGQLEQESCLDQWERSQ 121
Db 76 LTNDTQIL-----SKOYNETQNVFRAFTSAP-VDTYSRISLVPQNGWDYDN 124
QY 122 DVLSTIVTEWNLVCDWKAPLTSLFFVGLGSLFSGOLSDRFGKKNLFLVTMGHQT 181
Db 125 STYLSLVTETNVLCDQQAWEISTSPVGSFIGNCLFGVADKFGRRSFFVILTVLI 184
QY 182 GFSLQIFSKNFMFVFLVVLVGMQGISNYAAVFLGTEILGKSVRIIFSLGVCIFVAF 241
Db 185 VCGTASSPAKDIESFILLRFTTGLAPALFQIPFTICMEFGNKGSR-IFSGMLTSLFPGA 243
QY 242 GYNVLPFLAYFTRDRWMLLVALTMGVLCVALWVLPSPRWLSIQGFEAEVILIRKAA 301
Db 244 AMALLGVWAMFTRRQLTFFCNPAFAFYIYFFLPSPRWSVSGKWAADKOLKKIA 303
QY 302 KANG-----IVVPTIFDPSLQDSKKQ--QSHNLDLRLTWNIRMTVIMSLMWTISV 356
Db 304 KNGKSNVDVDELVDMSKKNHQAABEKEKRSNHTDLFKTPNLRRKTLIVYIWMVNAI 363
QY 357 GYFGLSLDPNHLGDFIVNCLFSAVMEVPAYVLAULLQYLPRRYSMATATFLGGSVILF 416
Db 364 IYNGLTANSLNPLVDYWSFIINGAVELPGYFVWVPLLCQACGRRTLAATMIVCGIGCVS 423
QY 417 MOLVPPDLYLATLVLMVCKGFTAFSAVMVYVTAELPTVVRNMGVGVSSPASLGSIL 476
Db 424 AMFMDPGYPLWVASAFIOKFGVGSFAVIYIFAGELPTVVRNMGVGVSSPASLGSIL 483
QY 477 SPYFVYGLYADRFPLMGSITLITAILTLPLPSFGTGLPDDTD 522
Db 484 APHIVNGLKTVKTLPLINGLMAISAGILTFELPTELGLAPLPTIE 529
RESULT 11
QYV6L6
ID Q9V6L6 PRELIMINARY; PRT; 561 AA.
AC Q9V6L6;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE CG4630 PROTEIN.
GN CG4630.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananidis P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H.C., Blasej R.G., Champs M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.L.G.,
RA Abrial J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harlowe D., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy K., Nusslein D.R., Pacle J.M.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusslein D.R., Pacle J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster";
RL Science 287:2185-2195(2000).
DR EMBL; AE003819; AAF58407.1; -.
DR FlyBase; FBgn0033809; CG4630.
DR InterPro; IPR001066; -.
DR Pfam; PF00083; sugar_tr. 1.
DR PIR; P00083; sugar_tr. 1.
SQ SEQUENCE 561 AA; 62041 MW; 573643481595788F CRC64;
Query Match 26.1%; Score 753.5; DB 5; Length 561;
Best Local Similarity 33.2%; Pred. No. 4.4e-40;
Matches 184; Conservative 120; Mismatches 200; Indels 51; Gaps 14;
QY 3 DYDEVTAFLEWGFQRLIFLFLSASIIPIGFTGLSSVFLIATPEHRCRVPDANLSSA- 61
Db 14 DFDITLVEIGFGRFORNYLLICLPVLAANSLSVFTAGSTYRCYPCDKLVDAE 73
QY 62 ---WRNHTVP-LRLDGRVPHSCRYRLATIANPSAL--GLEPGRDVLGQ----- 107
Db 74 YGANVSVAVPGSWSKRGHTPSTCERF-----VANGHLESSSDPWSAWPLDQCFAENFT 129
QY 108 LEQESCLDQWERSQDQVYLS---TIVTEWNLVLC-EDDKAPLITISLFFVGLGSLFISQGL 163
Db 130 TETERC-----NQFVYSSERTIVQQWGLQCPENLKNLAFVGTIFHAGLVGVGTALSGYL 183
QY 164 SDRGRKNV-LFVTMGQ-TGFSLQIFSKNFMFVFLVVLVGMQGISNYAAVFLGTEI 221
Db 184 ADYGRKHIFLCIVFVMTALGVA--QALSWDYISLFFALLNAVGTSGVYPLAFIIGVEM 241
QY 222 LKSKVRIIFSTLGVCIYFAGYVMLPLFAYFIRDRWMLLVALTMGVLCVALWVLPFISP 281
Db 242 VGPKRREM-SSIVLNYFYAVGEALLGL-SVFLPDWRQLQALSVPLPICVAYFWLWVPSV 299
QY 282 RWLISQGRFEAEVILIRKAANGVIVPSTIFDPSLQDSKKQSHN----- 330
Db 300 RWLLARNRREGAGVILIRRAAKVNRDRISVELMASFKQQLDAETQEDDVEGGLHVKDD 359
QY 331 -----ILDLRTNIRMTVIMSLMWTISVGYTGLSLDTPNLHGDFVNCFLSAMVEVP 385
Db 360 KIWLAVKEVARSHILMGRYAILLIWVNAIVYGLSNATSLSGNKNYFNALVCLVEIP 419
QY 386 AVYLAWLILLOVLPYRYSMATATFLGGSVLLFMQLVPPDLYLATVLMVGVKFGVTAAFSM 445
Db 420 GYSLAWLFLRGRVAVLUSGLLCSITCVASGFTVIGANMLVTVTLVGLGLTSSFAV 479
QY 446 VVVYTAELYPTVVRNMGVSVSTASRLGSIISPVVFLGAYDRFLPYILMGSILITAIL 505
Db 480 IYTFTEAMPTVIRSGGVGVMSVTFARFAGMLAPFVPLLASVYDPLPLLLFGLTSLVAGLL 539
QY 506 TILFSPESFGTPLPDT 520
Db 540 SLLLPETFNRLPDT 554
RESULT 12
ID Q9N0D4 PRELIMINARY; PRT; 554 AA.
AC Q9N0D4;
DT 01-OCT-2000 (TREMBlrel. 15, Created)

DT	01-OCT-2000 (TEMBLrel. 15, Last sequence update)
DE	01-MAR-2001 (TEMBLrel. 16, Last annotation update)
DT	PARTIAL OCT1 ORGANIC CATION TRANSPORTER, EXON 1 AND JOINED CDS.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX	NCBI_TaxId=9606;
RN	[1]
RP	SEQUENCE FROM N.A.
RA	Hayer M., Bonisch H., Bruss M.;
RT	"Molecular cloning, functional characterization and genomic
RT	organization of four alternatively spliced isoforms of the human
RT	organic cation transporter 1 (hOCT1/SUC22A1).";
RL	Ann. Hum. Genet. 63:473-482(1999).
[2]	
RP	SEQUENCE FROM N.A.
RA	Hayer M., Bonisch H., Bruss M.;
RT	"Molecular cloning, functional characterization and genomic
RT	organization of four alternatively spliced isoforms of the human
RT	organic cation transporter 1 (hOCT1/SUC22A1) (corrigendum).";
RL	Ann. Hum. Genet. 64:267-267(2000).
DR	EMBL; AJ243995; CAB95971.1; .
DR	EMBL; AJ243996; CAB95971.1; JOINED.
DR	EMBL; AJ276051; CAB95971.1; JOINED.
DR	EMBL; AJ276052; CAB95971.1; JOINED.
DR	EMBL; AJ276053; CAB95971.1; JOINED.
DR	EMBL; AJ245460; CAB95971.1; JOINED.
DR	EMBL; AJ243998; CAB95971.1; JOINED.
DR	EMBL; AJ243999; CAB95971.1; JOINED.
DR	EMBL; AJ244000; CAB95971.1; JOINED.
DR	InterPro: IPR001066; .
DR	Pfam: PF00083; sugar_tr; 1.
DR	PROSITE: PS00216; SUGAR_TRANSPORT_1; UNKNOWN_2.
DR	SEQUENCE 554 AA; 611169 MW; 9CA1A6G0FC2B2A96 CRC64;

Query Match 25.7%; Score 740; DB 4; Length 554;

Best Local Similarity, 35.9%; Pred. NO. 3.1e-39;
Matches 197; Conservative 86; Mismatches 215; Indels 50; Gaps 12;

QY	1	MRDYDEVTAFLGKGGPFOR---	LIFFILLSSAIIPNGFTGLSSVFLIATP	HEHRCVPPDAAAN	57	
Db	1	MPTVDDILEQVGESGWFOKAFLLCLLSAAFAF---	ICVGIVFLGFTPDHHCQSPGVAE	57		
QY	58	LSS--AWR-----	NHTVPLRLRDREVPVHSCRRYR---	LATIA-NFSALG	96	
Db	58	LSORCGWSPAEBELNYTV	PGLCPAGEAFLGCRQRYEVDWQNSALS	CVDFPLASLATNRGHL	117	
QY	97	LEBGRDVLGQLEQESCLDGWFSQDQVYI	STIVTENLVNLCEDDWKAPLTIISLFFVGVLLG	156		
Db	118	LGP-----	CODGWWY--DTPGSSIVTEFNLCADSKWLDL	FSCLNAGFLG	162	
QY	157	SFTSGQLSDRFGRKNVLFTVMGTGFS	FLQIFSKNFEMFVFLVLCMGQOISNVA	AAFV	216	
Db	163	SLGCVYFADRGFKLCLLT	TVLVNNAVSGVLMAFSPNYSMLFLRL	QGLGLKSGKNMAGYT	222	
QY	217	LQTEILGKSVRIIFSTLGVCFYAF-	GYMVLPLFAYFTRDWRMLLVALT	MPGVLCVAM	274	
Db	223	LITEFVSGSR---	RTVAIMQMAFTVGVALTGLAYALPHWRML	QLAVSLPTFLFLYY	279	
QY	275	WPIPESRWLLISOGREFEAEVIIR	KAAKANGIVPSTFIDPSELQDLSKK	QQOSHNLDL	334	
Db	280	WCVPESRWLLISQKRNTEAKI	MHDHTAQNGKLPADUKMLSLEDVTEK-	USPSPADL	337	
QY	335	LRTWIRWVTWISLMLNMTTIS	GVYFGLSLDTPNLHGDIIVANCFLS	AMVEPAYVYLA	394	
Db	338	FRTLRLKRTFILMYLFTDS	VLVYQGLIILHMGATSGNLYDL	FLYSALVEIPGAFIALITI	397	
QY	395	QYLPYRYSNATALEFLGGSV	LELMQVPPDLXYLATLVLMVGKFGV	TAAFSVMVYVYTAELY	454	
Db	398	DRVGRIYPMANSNLLAGAAC	LVMIFFISPDHLNLTIIICVGRMG	GITTAIOMICLVNAELY	457	
QY	455	PTVVRNMGVGVSTASRLGS	ILSPFYVY-LGAYDRFLPYILMG	SLITILAITLFI	PESEF	513

```
Db      458 PFYFVNLGVWVCSSLCDDGGIITFFIVFRREVQWLPLFLFAVLGLAAGVTLLLPETK 517  
               |||||::|| | : |: ::| | : | || | : || | | : || | |  
  
Qy      514 GTPLPDTI 521  
               I   | | : |  
  
Db       518 GVALPETM 525  
  
RESULT 13  
075751  
ID    O75751          PRELIMINARY;           PRT;         556 AA.  
AC     O75751; Q9UP02;  
DT     01-NOV-1998 (TReMBLeRel_08, Created)  
DT     01-MAR-1998 (TReMBLeRel_08, Last sequence update)  
DT     01-NOV-2001 (TReMBLeRel_16, Last annotation update)  
DE     ORGANIC CATION TRANSPORTER 3 (EXTRANEURONAL MONOAMINE TRANSPORTER)  
DE     (EMT) (SOLUTE CARRIER FAMILY 22, MEMBER 3).  
GN     SLC22A3 OR EMTH.  
OS     Homo sapiens (Human).  
OC     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC     Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX     NCBI_TaxId=9606;  
[1]  
RN     SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.  
RP     TISSUE=KIDNEY;  
RC     MEDLINE=99212254; PubMed=10196521;  
RA     Gruendemann D., Schechinger B., Rappold G.A., Schoemig E.;  
RT     "Molecular identification of the corticosterone-sensitive  
RL     extraneuronal catecholamine transporter." ;  
RL     Natl. Neurosci. 1:349-351(1998).  
[2]  
RN     SEQUENCE OF 513-556 PROM N.A., AND TISSUE SPECIFICITY.  
RP     MEDLINE=99134299; PubMed=933568;  
RA     Verhaegh S., Schweifer N., Barlow D.P., Zwart R.;  
RT     "Cloning of the mouse and human solute carrier 22a3 (SLC22a3/SCL22A3)"  
RI     identifies a conserved cluster of three organic cation transporters or  
RL     mouse chromosome 17 and human 6q26-q27.";  
RL     Genomics 55:209-218(1999).  
[3]  
RN     FUNCTION, AND TISSUE SPECIFICITY.  
RP     TISSUE=KIDNEY;  
RC     MEDLINE=20425388; PubMed=10966924;  
RA     Wu X., Huang W., Ganapathy M.E., Wang H., Kekuda R., Conway S.J.,  
RA     Leibach F.H., Ganapathy V.;  
RT     "Structure, function, and regional distribution of the organic cation  
RL     transporter OCT3 in the kidney";  
RL     Am. J. Physiol. 279:F449-F458(2000).  
CC     -1- FUNCTION: MEDIATES POTENTIAL-DEPENDENT TRANSPORT OF A VARIETY OF  
        ORGANIC CATIONS.  
CC     -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).  
CC     -1- TISSUE SPECIFICITY: EXPRESSED IN PLACENTA, SKELETAL MUSCLE,  
CC     PROSTATE, AORTA, LIVER, FETAL LUNG, SALIVARY GLAND, ADRENAL GLAND,  
CC     KIDNEY AND BRAIN CORTEX. NO EXPRESSION DETECTED IN SPLEEN.  
CC     -1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY. ORGANIC  
        CATION SUBFAMILY.  
DR     EMBL; AJ004417; CAAD04751.1; -.  
DR     EMBL; AF008749; AAA20977.1; -.  
DM     MIG; 604842; -.  
DR     InterPro; IPR001066; -.  
DR     Pfam; PF000083; sugar_trt; 1.  
DR     PROSITE; PS00216; SUGAR_TRANSPORT_1; 1.  
DR     PROSITE; PS00217; SUGAR_TRANSPORT_2; FALSE_NEG.  
KW     Transport; Ion transport; transmembrane; Glycoprotein.  
FT     TRANSMEM            21              41  
FT     TRANSMEM            182             202  
FT     TRANSMEM            241             261  
FT     TRANSMEM            269             289  
FT     TRANSMEM            381             401  
FT     TRANSMEM            468             488  
FT     TRANSMEM            498             518  
FT     CARBOHYD             72              72  
FT     CARBOHYD             99              99  
FT     CARBOHYD            119             119
```

FT CARBOHYD 322 322 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 556 AA; 61279 MW; C3CA2D7DD21C658 CRC64;

Query Match 25.6%; Score 738.5; DB 4; Length 556;
Best Local Similarity 33.0%; Pred. No. 3.9e-39;
Matches 189; Conservative 117; Mismatches 214; Indels 53; Gaps 18;

QY 1 MRDYDEVTAFLGEGWGFQRLIFELLSASIIIPNGFTGLSVFLIATPEHR-CRVPDAANLS 59
DB 1 MPFDEALQVGEFGFRFRRVLLCLTGVTAFILFVGVGTGTPDHYWCRGSPSAALA 60
QY 60 S-AWR-----NHTVPLRLRDGVEPH-----SCRRVRLATIANFSA-----LGLE 98
DB 61 ERGCWSEPEWNTAP--ASRGPEPERGRGCRY--LLEAANDSASATLSADPLAFL 117
QY 99 PGRDVLGQLEQESCLDGNFESQDVYLSITVTENWLVCEDDKAPLITISLFFVGVLLGSF 158
DB 118 PNRSAFL-----VPCRGGWRYAQ--AHSTIVSEFDLVCVNWMLDLTQAILNLGLTGA 170
QY 159 ISQLSDRGKRWLVFTVMQMGTFSELOI-PSKNEFMFVFLVLMGQISNYVAAPVL 217
DB 171 TLGYAADRYGR-IVYILLSCLVGVTVGVVAFAPNPVFVIFRLOGVFGKGTWMTCYVI 229
QY 218 GTILGKSVRIIFSTLGVCII--FYAFGYMVLPLFAFYFIRDRWMLVALTMPGVLCVALMW 275
DB 230 VTEIVSKQRI---VGIWQMFTEGIIILGIAFYFNWQIQIATILPSEFLYYW 286
QY 276 FIPESRWLISQGRFEAEVIRKAAKANGIVVPSTIFDPSELODSSKKQOOSHNLIDL 335
DB 287 VVPESPRWLITRKKGKALQILRIAKCKNGKYLSSNY---SEI-TVTDEEVNPSFLDIV 342
QY 336 RTNIRNVTIMSLMWTISVGVFGLSLDTPNLHGDI FVNCFLSAMVEVPAYVLAWLLQ 395
DB 343 RTPQMRKCTILMFAMFTSAVYVQGLVMRLGIIGNLYIDFISGVVLPGLALLITIE 402
QY 396 YLPFRYSMATALFLGSGVLLFMQLVPPDLYLATVLMVKGCVTAFAFSMVVYTAELXP 455
DB 403 RLGRRLPFAASNIVAGVACLVATFLPEGTANLRTVATLGRGITWAFIVLVNSELXP 462
QY 456 TVVRNMGVSVSTASRLSGILSPFYVY-IGAYDRFLPYILMGSLTILTAILTLFLPESG 514
DB 463 TTLNFGVSCSLCDFGGLIAPFLFRLAAVWLEPLIIFGLIASICGGLVMLLPETKG 522
QY 515 TPLPDTIDQLRV-----KGMKHKRTP---SH 538
DB 523 IALPETVDDVEKLGSHPSCCKGRNKKTPVSRSH 555

RESULT 14
O15395 PRELIMINARY; PRT; 554 AA.
AC O15395;
DT 01-JAN-1998 (Tremblrel. 05, Created)
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
DE 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE ORGANIC CATION TRANSPORTER 1.
GN HOCT1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97330822; PubMed=9187257;
RA Zhang L., Dresser M.J., Gray A.T., Yost S.C., Terashita S.,
RA Giacomini K.M.;
RT "Cloning and functional expression of a human liver organic cation
RT transporter";
RL Mol. Pharmacol. 51:913-921(1997).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -!- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.
DR EMBL; U77086; AAB67703.1; -.

DR InterPro; IPR001066; -.
DR Pfam; PF00083; sugar_tr; 1.
DR PROSITE; PS00216; SUGAR_TRANSPORT_1; UNKNOWN_2.
KW Transmembrane.
SQ SEQUENCE 554 AA; 61153 MW; 55206B897DE32202 CRC64;

Query Match 25.6%; Score 738; DB 4; Length 554;
Best Local Similarity 35.9%; Pred. No. 4.2e-39;
Matches 197; Conservative 86; Mismatches 215; Indels 50; Gaps 12;

QY 1 MRDYDEVTAFLGEGWGFQRLIFELLSASIIIPNGFTGLSVFLIATPEHRCKRVPDAAN 57
DB 1 MPTVDDILQEVGSGWGFQKQAFILCLLSAFAF---ICVGIPLFGTTPDHCCSPGVAE 57
QY 58 LSS-AWR-----NHTVPLRLRDGVEPHSCRRY-----LATIA-NFSAIG 96
DB 58 LSQRCGSPAEELNVTVPGLGPAGEAFILGOCRRYEVDMNQSSALSCVDPLASLATNRSHLP 117
QY 97 LEPRGVDLQLEQESCLDGNFESQDVYLSITVTENWLVCEDDKAPLITISLFFVGVLLG 156
DB 118 LGP-----CODGWY--DTPGSSIVTEFNLCVADSKLDFQSLNAGFLG 162
QY 157 SFISQSLSDRGKRWLVFTVMQMGTFSELOI FSKNEFMFVFLVLMGQISNYVAAFV 216
DB 163 SLGVGYFADRGKRLCLLGTVLVNAVSGVLMFSPNYMSMLLFRLLQGLVSKGNMAGYT 222
QY 217 LGTEILGKSVRIIFSTLGVCIFYAF--GYMVLPLFAFYFIRDRWMLVALTMPGVLCVALW 274
DB 223 LITEFVGVSGSR---RTVAIMYQMAFTGLVALTGLAYALPHRWLQLAVALSPLTFELYY 279
QY 275 WFIPESRWLISQGRFEAEVIRKAAKANGIVVPSTIFDPSELODSSKKQOOSHNLIDL 334
DB 280 WCVPESRWLISQGRFEAEVIRKAAKANGIVVPSTIFDPSELODSSKKQOOSHNLIDL 337
QY 335 LRTNIRNVTIMSLMWTISVGVFGLSLDTPNLHGDI FVNCFLSAMVEVPAYVLAWLL 394
DB 338 FRTPLRKRRTFLMWFTDSVLYQGLILHMGATSGNLYLDLYLSALVEIPGAFIALITI 397
QY 395 QYLPFRYSMATALFLGSGVLLFMQLVPPDLYLATVLMVKGCVTAFAFSMVVYTAELXP 454
DB 398 DRVGRYIPMANSNLAGAACLVMI FISPDLHMLNIIIMCVGRMGITTAIOMICLVNAELY 457
QY 455 PTVRNMGVSVSTASRLSGILSPFYVY-IGAYDRFLPYILMGSLTILTAILTLFLPESF 513
DB 458 PTFVRNMGVSVSTASRLSGILSPFYVY-IGAYDRFLPYILMGSLTILTAILTLFLPESF 517
QY 514 GTPLPDTI 521
DB 518 GVALPETM 525

RESULT 15
O15245 PRELIMINARY; PRT; 554 AA.
AC O15245;
DT 01-JAN-1998 (Tremblrel. 05, Created)
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE ORGANIC CATION TRANSPORTER.
DE Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97405886; PubMed=9260930;
RA Gorboulev V.G., Ulzheimer J.C., Akhondova A.;
RT "Cloning and characterization of two human polyspecific organic cation
RT transporters";
RL DNA Cell Biol. 16:871-881(1997).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).

```
CC -I- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.
DR EMBL; X98332; CAA6977.1; -.
DR InterPro; IPR001066; -.
DR Pfam; PF00083; sugar tr. 1.
DR PROSITE; PS00216; SUGAR_TRANSPORT_1; UNKNOWN_2.
KW Ion transport; Transmembrane.
SQ SEQUENCE 554 AA; 61187 MW; B16E200852CC4000 CRC64;

Query Match      25.5%; Score 734; DB 4; Length 554;
Best Local Similarity 35.8%; Pred. No. 7.5e-39;
Matches 196; Conservative 86; Mismatches 216; Indels 50; Gaps 12;

QY 1 MRDYDEVTAFLGEMGPQR---LIFFLLSASIIIPNGFTGLSSVFLIATPEHRCRVPDAAAN 57
   |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1 MPTVDDILEQVGSGWFKQAFILICLLSAAPAP---ICVGIVFLGFTPDHHCQSPGVAE 57

QY 58 LSS--AWR-----NHTVPLRLRGREVPHSCRRY-----LATIA-NFSALG 96
   |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 58 LSORCGWSPAEELNYTPVGLGPAGEAFLGCRRYEYDWNQSAALSCVDPLASLATNRSLP 117

QY 97 LEPRGDVLDLQLEQESCLDGEESQDVYLTSTIVTENLVCEDDWKAPLITISLFFVGVLLG 156
   |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 118 LGP-----CQDGWY--DTPGSSIVTEFNLCADSWKLDLFQSCINAGFFEG 162

QY 157 SFISGQLSDRGRKNVLFVTMGQTFGSLQIFSKNFEMFVVLVFLVGMGQISNYAAEV 216
   |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 163 SLGVGYFADRFGRKLCILGTGLVNAVSGVLMAPSPNYSMLLFRLLQGLVSKGNMAGYT 222

QY 217 LGTEILGKSVRIIFSFLGVCIFYAF--GYMVLPLFAYFIRDRHMLLVALTMPGVLCAVM 274
   |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 223 LITEFVGSSR---RIVAIMYQMAFTVGLVALTGLAYALPHRWLQOLAVSLPTFLFLYY 279

QY 275 WFIPESPRWLISQRFEEAEVIRKAAKANGIVWPSTIEDPSELQDLSSKKQOOSHNLIDL 334
   |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 280 WCVPESPRWLLSQKRNTEAIKINDHIAQKNGKLPADLKLMSLEEDVTEK--LSPSFADL 337

QY 335 LRTWNIRMTVIMSILMWTISVGYFGLSLDTPNLHGDIIVNCFLSAMVEVPAYVLAWLL 394
   |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 338 FRTPRLRKRTFILMYLWFTDSVLYQGLIILHMGATSGNLYLDFLYSALVEIPGAFIALITI 397

QY 395 QYLPRRYSNATALFLGGSVLLFMOLVPPDLYLYLATVLMVWKGFGVTAAPSMVYVYTAELY 454
   |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 398 DRVGRIYPMAMSNLLAGACLWIFISPDHLWLNIIIMCVGRMGITIAIQMICLVNAELY 457

QY 455 PTVRNMGVGSVTSARLGSILSPYEVY-LGAYDRFLPYILMGSLTILTAITLFLPESF 513
   |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 458 PTEVRNLGVNMCSSLCDIGGIITPTFIVRLREVWQALPLILFAVLGLLAAGVYLLLPETK 517

QY 514 GTPLPDTI 521
   |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 518 GVALPETM 525
```

Search completed: August 16, 2001, 14:03:36
Job time: 308 sec